

Similarity	100.0%	Pred	No	0
------------	--------	------	----	---

	Matches	2390	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	AGAGGAGACGGACGTTTGAGAGAACGAGAGAGAAAGAGAGAAATGGCGTTCACGGATTA	60							
Db	1	AGAGGGAGACGGACGTTTGAGAGAACGAGAGAGAAAGAGAAATGGCGTTCACGGATTA	60							
QY	61	CAGTACCTTATAGCCACGTGACGGGCGACGAGGGGCTACAGTCTTACACCGGCCAGCCAC	120							
Db	61	CAGTACCTTATAGCCACGTGACGGGCGACGAGGGGCTACAGTCTTACACCGGCCAGCCAC	120							
QY	121	TCAAGATATGACAGACCAACCCAGGCAATATGGCAACAAAGCTATGGAACTTATGACAA	180							
Db	121	TCAAGATATGACAGACCAACCCAGGCAATATGGCAACAAAGCTATGGAACTTATGACAA	180							
QY	181	GCCCACTGATGTCACTATACCCAGGCTCAGACCACTGACCACTATGGGCAAGCCGCTTA	240							
Db	181	GCCCACTGATGTCACTATACCCAGGCTCAGACCACTGACCACTATGGGCAAGCCGCTTA	240							
QY	241	TGCAACTTCTTATGAGACAGCCTCCACCTGGTTATACTCACTGACCTGCCCCAGGCACTA	300							
Db	241	TGCAACTTCTTATGAGACAGCCTCCACCTGGTTATACTCACTGACCTGCCCCAGGCACTA	300							
QY	301	CAGCCAGCCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCAACACTGCTACAGTCAAC	360							
Db	301	CAGCCAGCCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCAACACTGCTACAGTCAAC	360							
QY	361	CACCAACCCAGGCTCTTATGACGCTCACTGCAATATGGCACTACGCTGCTTATCCAGC	420							
Db	361	CACCAACCCAGGCTCTTATGACGCTCACTGCAATATGGCACTACGCTGCTTATCCAGC	420							
QY	421	CTATGGGCAAGGACAGACGACCACTGACCTTCAACAAACCGAGATGGAAACAAGCCAC	480							
Db	421	CTATGGGCAAGGACAGACGACCACTGACCTTCAACAAACCGAGATGGAAACAAGCCAC	480							
QY	481	TGAGACTAGTCAACCTCAATATAGCAACAAGGGGTTTACAACAGCCACAGCTAGGATATG	540							
Db	481	TGAGACTAGTCAACCTCAATATAGCAACAAGGGGTTTACAACAGCCACAGCTAGGATATG	540							
QY	541	ACAGAGTAACTTACAGTTATCCCAAGGTAAGCTGGAGACTACCCCATGACCAAGTCACTGC	600							
Db	541	ACAGAGTAACTTACAGTTATCCCAAGGTAAGCTGGAGACTACCCCATGACCAAGTCACTGC	600							
QY	601	ACCTTCATCTTACCCCTCCTTACCAAGTATTCCTCTTACAAGCCGACTAGTTATGATCAAG	660							
Db	601	ACCTTCATCTTACCCCTCCTTACCAAGTATTCCTCTTACAAGCCGACTAGTTATGATCAAG	660							
QY	661	CAGTTACTCTCAGCGAAGACCTTATGGCAACCGACAGCTATATGACAGCAGAGTACGTA	720							
Db	661	CAGTTACTCTCAGCGAAGACCTTATGGCAACCGACAGCTATATGACAGCAGAGTACGTA	720							
QY	721	TGGTCAACAAAGCACTATATGGGCAAGCCTCCACTATAGTTAACCAACCCAACTGATC	780							
Db	721	TGGTCAACAAAGCACTATATGGGCAAGCCTCCACTATAGTTAACCAACCCAACTGATC	780							
QY	781	CTACAGCCAAAGTCCAAATATATAGCCAAACAGACGACGCTATCGGGCAGCAGAGTTC	840							
Db	781	CTACAGCCAAAGTCCAAATATATAGCCAAACAGACGACGCTATCGGGCAGCAGAGTTC	840							
QY	841	ATTCCGACAGGACCAACCCAGTACGATGGGATGTTATATGGGACAGAGTCTGGAGATTTTC	900							
Db	841	ATTCCGACAGGACCAACCCAGTACGATGGGATGTTATATGGGACAGAGTCTGGAGATTTTC	900							
QY	901	CGGACACAGGAGAAACCGGACATATGATGGCCCTGATTAACCGGGGCAAGGGAGAGGGGG	960							
Db	901	CGGACACAGGAGAAACCGGACATATGATGGCCCTGATTAACCGGGGCAAGGGAGAGGGGG	960							
QY	961	ATTGATATGTGAGAGGCAATGACGAGAGTGGGCGGGGAGAGAGACCGGCTGATATGGGAG	1020							
Db	961	ATTGATATGTGAGAGGCAATGACGAGAGTGGGCGGGGAGAGAGACCGGCTGATATGGGAG	1020							
QY	1021	CGCTGGAGAGGAGGATGGCTTCAATTAACCTGTGTGAGACCAATGATGTAAGGACCAAGTCT	1080							
Db	1021	CGCTGGAGAGGAGGATGGCTTCAATTAACCTGTGTGAGACCAATGATGTAAGGACCAAGTCT	1080							

QY	1081	TGATCTAGAGCCCACTCTGTATATCCAGATGAAGA	CTCTGACAAACAGTSCAAATTATATGACA	1140
Db	1081	TGATCTAGAGCCCACTCTGTATATCCAGATGAAGA	CTCTGACAAACAGTSCAAATTATATGACA	1140
QY	1141	AGGATTTAAATGACAGTGTGA	CTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGAGT	1200
Db	1141	AGGATTTAAATGACAGTGTGA	CTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGAGT	1200
QY	1201	TGTTAAGATGACAAAGAGAACTBGGGCAACC	CATGATGATCAATCTACCTGGCAAGAGAAAC	1260
Db	1201	TGTTAAGATGACAAAGAGAACTBGGGCAACC	CATGATGATCAATCTACCTGGCAAGAGAAAC	1260
QY	1261	AGGAAAAGCCCAAGGCGATGCCA	CAGTGTCTTATGAAAGCCCACTGCGCAAGGCTGC	1320
Db	1261	AGGAAAAGCCCAAGGCGATGCCA	CAGTGTCTTATGAAAGCCCACTGCGCAAGGCTGC	1320
QY	1321	CGTGGAAATGGTTTGATGTGGAAAGAAATTTT	CAAGGAGCGAACTTAAATCTTCTCTTGGTCTCG	1380
Db	1321	CGTGGAAATGGTTTGATGTGGAAAGAAATTTT	CAAGGAGCGAACTTAAATCTTCTCTTGGTCTCG	1380
QY	1381	GAAAGAAAGCTTCCATATGAACGTATGCGGGGT	TGGTCTGCCACCCCGTGAAGGCGAGAGGCAT	1440
Db	1381	GAAAGAAAGCTTCCATATGAACGTATGCGGGGT	TGGTCTGCCACCCCGTGAAGGCGAGAGGCAT	1440
QY	1441	GCCACCAACCACTCCGTGTGAGAGTCC	AGAGAGGCCAAGAGGCTCTGGGGGAAACCATAGGTTGC	1500
Db	1441	GCCACCAACCACTCCGTGTGAGAGTCC	AGAGAGGCCAAGAGGCTCTGGGGGAAACCATAGGTTGC	1500
QY	1501	CATGGGAGGCGCGTGGAGAGAAATGAGAGAG	CTTCCCTCAAGAGAGACCCCGGGGTTTCCCG	1560
Db	1501	CATGGGAGGCGCGTGGAGAGAAATGAGAGAG	CTTCCCTCAAGAGAGACCCCGGGGTTTCCCG	1560
QY	1561	AGGGAACCCCTCTGTGAGAGAGAAACGT	CCAGACCCGAGCTGGAGACTTGGCAGTGTCCCAA	1620
Db	1561	AGGGAACCCCTCTGTGAGAGAGAAACGT	CCAGACCCGAGCTGGAGACTTGGCAGTGTCCCAA	1620
QY	1621	TCCGGGTTTGTGGAAACCA	GAACTTCCGCTGTGAGAAACAGATGTGCAACCAAGTAAAGGCCCC	1680
Db	1621	TCCGGGTTTGTGGAAACCA	GAACTTCCGCTGTGAGAAACAGATGTGCAACCAAGTAAAGGCCCC	1680
QY	1681	AAAAGCTGGAAGGCTTCTCTCCGCCAC	CCCTTTCGCCGCCCGGGGTGTGATCGTGGCAGAGG	1740
Db	1681	AAAAGCTGGAAGGCTTCTCTCCGCCAC	CCCTTTCGCCGCCCGGGGTGTGATCGTGGCAGAGG	1740
QY	1741	TGCGCTCTGTGTGGCATGCGGGGAGGAA	AGGTGGCTCATGTGATGTGTGTCTCCGGTGG	1800
Db	1741	TGCGCTCTGTGTGGCATGCGGGGAGGAA	AGGTGGCTCATGTGATGTGTGTCTCCGGTGG	1800
QY	1801	AAATGTTCAAGAGTGGCCGTGTGTGAGACA	GAGAGTGGCTTCCGTGTGTGGCCGGGGCATGGA	1860
Db	1801	AAATGTTCAAGAGTGGCCGTGTGTGTGAGACA	GAGAGTGGCTTCCGTGTGTGGCCGGGGCATGGA	1860
QY	1861	CCGAGGTGGCTTTTGTGTGTGAGAGAA	CCAGAGTGGCTTGGGGGGGCCCTCTGGACCTTTGAT	1920
Db	1861	CCGAGGTGGCTTTTGTGTGTGAGAGAA	CCAGAGTGGCTTGGGGGGGCCCTCTGGACCTTTGAT	1920
QY	1921	GGAACCAATGTGGAGAGAAAGAGAGAGAC	GTGTGAGACCTTGGAAAAATTGTGATAAAGCGA	1980
Db	1921	GGAACCAATGTGGAGAGAAAGAGAGAGAC	GTGTGAGACCTTGGAAAAATTGTGATAAAGCGA	1980
QY	1981	GCAACCGTCAGAGAGCGCAGAGATCGG	CCCTACTATATGCAAGAGACCCCGCAGAGTGCATTT	2040
Db	1981	GCAACCGTCAGAGAGCGCAGAGATCGG	CCCTACTATATGCAAGAGACCCCGCAGAGTGCATTT	2040
QY	2041	GACTACCAAGATTTATTTTTTAAAC	CAAGAAATGTTTTTAAATTTAATTTCAATTTTATA	2100
Db	2041	GACTACCAAGATTTATTTTTTAAAC	CAAGAAATGTTTTTAAATTTAATTTCAATTTTATA	2100
QY	2101	ATGTGTGGCCAAACATTTATGATTTAT	TTCCTTGTCTGTACTTTAGATATTTTTCACCAATTTGT	2160
Db	2101	ATGTGTGGCCAAACATTTATGATTTAT	TTCCTTGTCTGTACTTTAGATATTTTTCACCAATTTGT	2160



OY		2161	CAGGAAACATTTAAAACAAGTAAAAAGTAAGTCGTGGGAGTTTTTTTTCTCCTCTTTT	2220
Db		2161	GAGGAACAATTAAAACAAGTTAAAAAGTAGTGCTGGAGATTTTTTTTTCTTCCTCTTTT	2220
OY		2221	AAAAATGGTGTGTTAAGAATTAAACAATGGGAAACCCCTTGAGCATCTCGATATCAT	2280
Db		2221	AAAAATGGTGTGTTAAGAATTAAACAATGGGAAACCCCTTGAGCATCTCGATATCAT	2280
OY		2281	GTGGAAACCAGAGGGCCTCTTAACTGTAAACAATGTTCAATGTTGATGTTTTTTTTT	2340
Db		2281	GTGGAAACCAGAGGGCCTCTTAACTGTAAACAATGTTCAATGTTGATGTTTTTTTTT	2340
OY		2341	TTTTTTTAAATATAAATGCCAAATGTTTAAATATAAAAAAAAAAAAAAAAAAAAA	2390
Db		2341	TTTTTTTAAATATAAATGCCAAATGTTTAAATATAAAAAAAAAAAAAAAAAAAAA	2390
RESULT 2				
AX411125				
LOCUS	AX411125	2390 bp	DNA	linear PAT 14-JUN-2002
DEFINITION	Sequence 3772 from Patent WO0229103.			
ACCESSION	AX411125			
VERSION	AX411125.1 GI:21443830			
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Alvarez, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G. Gene expression profiles in liver cancer Patent: NO 0229103-A 3772 11-APR-2002;			
AUTHORS				
TITLE				
JOURNAL	GENE LOGIC INC (US) Location/Qualifiers 1..2390 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606" /note="EMBL/GenBank Accession No. X66899"			
FEATURES				
source				
ORIGIN				
Query Match	99.9%; Score 2388.4; DB 6; Length 2390;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 2389; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			
OY		1	AGAGGAGACGCGACTGTAGAGAAACGAGAGGAAAGAGAAATGGCGTCCACGATTA	60
Db		1	AGAGGAGACGCGACTGTAGAGAAACGAGAGGAAAGAGAAATGGCGTCCACGATTA	60
OY		61	CAGTACCTATAGCCAAAGTGCAGCGCGACAGGGGCTAACGTGCTTACACCGCCACGCCAC	120
Db		61	CAGTACCTATAGCCAAAGTGCAGCGCGACAGGGGCTAACGTGCTTACACCGCCACGCCAC	120
OY		121	TCAAAGATATGACAGCACCCACCGAGCATATGGGCAACAAAGCTATGGAACTTATGAC	180
Db		121	TCAAAGATATGACAGCACCCACCGAGCATATGGGCAACAAAGCTATGGAACTTATGAC	180
OY		181	GCCCATGATGTCACTATACCCAGGCTCAGACCACTGCAACTATATGGGACGACCGCTA	240
Db		181	GCCCATGATGTCACTATACCCAGGCTCAGACCACTGCAACTATATGGGACGACCGCTA	240
OY		241	TGCAACTCTTATGGACAGCCTCCACATGGTTATCTACTGCAACTGCCCCCAGGACATA	300
Db		241	TGCAACTCTTATGGACAGCCTCCACATGGTTATCTACTGCAACTGCCCCCAGGACATA	300
OY		301	CAGCCAGCCTGTCCAGGGGTATGGCACTGTGCTTATGATATACCACTGCTACAGTCAC	360
Db		301	CAGCCAGCCTGTCCAGGGGTATGGCACTGTGCTTATGATATACCACTGCTACAGTCAC	360
OY		361	CACCAACCCAGGCTCTTATGCAAGCTCACTTGCATATGGCACTGACCTGCTTATCCAGC	420
Db		361	CACCAACCCAGGCTCTTATGCAAGCTCACTTGCATATGGCACTGACCTGCTTATCCAGC	420

QY	421	CTATGAGGAGAGGAGCCAGGACCTGACCTTACAAAGACCGAGAGTGGAAACAAAGCCAC	480
Db	421	CTATGAGGAGAGGAGCCAGGACCTGACCTTACAAAGACCGAGAGTGGAAACAAAGCCAC	480
QY	481	TGAGACTGATCAACTCAATCTAGACAGAGGGGGTTACAAACAGGCCCTAGAGTTATGG	540
Db	481	TGAGACTGATCAACTCAATCTAGACAGAGGGGGTTACAAACAGGCCCTAGAGTTATGG	540
QY	541	ACAGAGTAACATCAAGTTATCCCAAGGTACTCTGGAGCTACCCCATGACGACATGAC	600
Db	541	ACAGAGTAACATCAAGTTATCCCAAGGTACTCTGGAGCTACCCCATGACGACATGAC	600
QY	601	ACCTTCATCTCACTCTTACAGCTATTCCTTACACAGCCGACTATGATCAAG	660
Db	601	ACCTTCATCTCACTCTTACAGCTATTCCTTACACAGCCGACTATGATCAAG	660
QY	661	CAGTTACTCTCAGAGAAACACTTATGGGCAACGAGAGCTATGACAGAGATGACTA	720
Db	661	CAGTTACTCTCAGAGAAACACTTATGGGCAACGAGAGCTATGACAGAGATGACTA	720
QY	721	TGCTCAACAAAGACAGCTATGGGAGAGAGCTCCCACTAGTTACCCCAACCTGAGATC	780
Db	721	TGCTCAACAAAGACAGCTATGGGAGAGAGCTCCCACTAGTTACCCCAACCTGAGATC	780
QY	781	CTACAGCCCAAGCTCCAACTATATAGCCAAAGAGCAGAGCTACGGGAGCAGAGTTTC	840
Db	781	CTACAGCCCAAGCTCCAACTATATAGCCAAAGAGCAGAGCTACGGGAGCAGAGTTTC	840
QY	841	ATTCCGAGAGAGCAACCCCACTAGCATTAGGGGTTATGGGAGAGATCTGGAGATTTTC	900
Db	841	ATTCCGAGAGAGCAACCCCACTAGCATTAGGGGTTATGGGAGAGATCTGGAGATTTTC	900
QY	901	CGGACCAAGAGAGAACCGGAGCATGATGAGGCTCTGATTAACGGGGAGAGGGAGAGGGG	960
Db	901	CGGACCAAGAGAGAACCGGAGCATGATGAGGCTCTGATTAACGGGGAGAGGGAGAGGGG	960
QY	961	ATTGATCTGAGAGGATGACAGAGGTGGGCGGGAGAGAGACGCGGTGAAATGGGAG	1020
Db	961	ATTGATCTGAGAGGATGACAGAGGTGGGCGGGAGAGAGACGCGGTGAAATGGGAG	1020
QY	1021	CGCTGAGAGAGGAGGTGGCTTCAATTAAGCTCTGTGAGCCCATGATGAAAGCAAGATCT	1080
Db	1021	CGCTGAGAGAGGAGGTGGCTTCAATTAAGCTCTGTGAGCCCATGATGAAAGCAAGATCT	1080
QY	1081	TGATCTAGGGCCCACTGTAGATCCAGATGAAGACTCTGACAAAGTGCAATTTATGTCA	1140
Db	1081	TGATCTAGGGCCCTCTGTAGATCCAGATGAAGACTCTGACAAAGTGCAATTTATGTCA	1140
QY	1141	AGGATTAATAGACAGTGTGACTCTAGATGATCTGGACAGACTTCTTTAAGCAGTGGGGT	1200
Db	1141	AGGATTAATAGACAGTGTGACTCTAGATGATCTGGACAGACTTCTTTAAGCAGTGGGGT	1200
QY	1201	TGTTAAGATGAACAAGAACTGGGCAACCATGATCCAATCTTACCTGAGCAAGAAAC	1260
Db	1201	TGTTAAGATGAACAAGAACTGGGCAACCATGATCCAATCTTACCTGAGCAAGAAAC	1260
QY	1261	AGGAAAGCCAAAGGCAATGCCAGAGTGTCTTAAAGACCCCACTGCAAGGCTGC	1320
Db	1261	AGGAAAGCCAAAGGCAATGCCAGAGTGTCTTAAAGACCCCACTGCAAGGCTGC	1320
QY	1321	CGTGAATGTGTTGATGGAAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTGC	1380
Db	1321	CGTGAATGTGTTGATGGAAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTGC	1380
QY	1381	GAAAGAGCTTCAATGAACAGTATGCGGGGTGTCTGCACCCGTGAGGCAAGAGCAT	1440
Db	1381	GAAAGAGCTTCAATGAACAGTATGCGGGGTGTCTGCACCCGTGAGGCAAGAGCAT	1440
QY	1441	GCACACCAACTCGGTGAGGTCCAGGAGGCGCCAGAGAGTCTTGAGGGAGCCCATGGGGTGC	1500
Db	1441	GCACACCAACTCGGTGAGGTCCAGGAGGCGCCAGAGAGTCTTGAGGGAGCCCATGGGGTGC	1500
QY	1501	CATGGAGGCGCTGAGAGAGATAGAGAGGCTTCCCTCAAGAGACCCCGGGGTTCCGC	1560



181 GCCCATGATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTTA 240  
181 GCCCATGATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTTA 240  
241 TGCATCTTCTTATGGAAGAGCTCCCACTGGTATTAATCTCAATGCTGCTCCCGGAGGATA 300  
241 TGCATCTTCTTATGGAAGAGCTCCCACTGGTATTAATCTCAATGCTGCTCCCGGAGGATA 300  
301 CAGCCAGCTGTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTCAAGTCAAC 360  
301 CAGCCAGCTGTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTCAAGTCAAC 360  
361 CACCAACCCAGGCTCTCTATGCAAGTCTGCAATATGGCACTGCAAGCTCTCTATATCAGC 420  
361 CACCAACCCAGGCTCTCTATGCAAGTCTGCAATATGGCACTGCAAGCTCTCTATATCAGC 420  
421 CTATGGGCAAGCCAGCAAGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 480  
421 CTATGGGCAAGCCAGCAAGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 480  
481 TGAAGTATGTCATCTCAATCTAGCAAGGGGGTTTACACAGCCCAAGCTAGGATATGG 540  
481 TGAAGTATGTCATCTCAATCTAGCAAGGGGGTTTACACAGCCCAAGCTAGGATATGG 540  
541 ACAGAGTATCTACAGTTATCCCAAGGTAAGCTGGAGGCTACCCCAAGCTAGCTAGCTGC 600  
541 ACAGAGTATCTACAGTTATCCCAAGGTAAGCTGGAGGCTACCCCAAGCTAGCTAGCTGC 600  
601 ACCCTCATCTCAACCTCTCTACAGCTATTCCTCTACACAGCCCACTAGTATATGATCAAG 660  
601 ACCCTCATCTCAACCTCTCTACAGCTATTCCTCTACACAGCCCACTAGTATATGATCAAG 660  
661 CAGTTACTCTCAGAGAACTATATGGGCAAGCCAGAGCTATGACAGAGAGTACTA 720  
661 CAGTTACTCTCAGAGAACTATATGGGCAAGCCAGAGCTATGACAGAGAGTACTA 720  
721 TGGTCAACAAAGAGCTATATGGGCAAGCCAGAGCTATGACAGAGAGTACTA 780  
721 TGGTCAACAAAGAGCTATATGGGCAAGCCAGAGCTATGACAGAGAGTACTA 780  
781 CTACAGCCAGAGCTCAAGTCAATATATAGCAACAGAGAGAGCTACGGGAGAGAGTTC 840  
781 CTACAGCCAGAGCTCAAGTCAATATATAGCAACAGAGAGAGCTACGGGAGAGAGTTC 840  
841 ATTCCGACAGAGCAACCCAGTATGAGTGGTATATGGGCAAGAGTCTGAGAGATTTTC 900  
841 ATTCCGACAGAGCAACCCAGTATGAGTGGTATATGGGCAAGAGTCTGAGAGATTTTC 900  
901 CGGACAGAGAGAACTGGAGCACTGAGTGGCTGATATACCGGGGAGGGGAAAGGGGG 960  
901 CGGACAGAGAGAACTGGAGCACTGAGTGGCTGATATACCGGGGAGGGGAAAGGGGG 960  
961 ATTGATCGTGGAGGCACTGAGAGAGTGGGCGGGGAGAGAGCGGGTGGAAATGGGAG 1020  
961 ATTGATCGTGGAGGCACTGAGAGAGTGGGCGGGGAGAGAGCGGGTGGAAATGGGAG 1020  
1021 CGCTGGAGAGCGAGGTGGCTTCAATAGCTGGTGGAGCCATGGATGAAAGCAAGATCT 1080  
1021 CGCTGGAGAGCGAGGTGGCTTCAATAGCTGGTGGAGCCATGGATGAAAGCAAGATCT 1080  
1081 TGAATCTAGGCGCCACTCTATGATCCAGATGAAAGCTGCAACAGTGCATTTATATGACA 1140  
1081 TGAATCTAGGCGCCACTCTATGATCCAGATGAAAGCTGCAACAGTGCATTTATATGACA 1140  
1141 AGGATTAATGAGAGCTGATCTGATATGATCTGGGAGCTTCTTTAAGAGAGTGGGGT 1200  
1141 AGGATTAATGAGAGCTGATCTGATATGATCTGGGAGCTTCTTTAAGAGAGTGGGGT 1200  
1201 TGTTAAGATGAAAG 1260  
1201 TGTTAAGATGAAAG 1260

1261 AGGAAAGCCCAAG 1320  
1261 AGGAAAGCCCAAG 1320  
1321 CGTGAATGGTGTGATGAGGAAAGATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
1321 CGTGAATGGTGTGATGAGGAAAGATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
1381 GAAAGAGCTCTCAATGAAAGATATGCGGGGTGCTGCAACCCGCTGAGGGAGAGAGCAT 1440  
1381 GAAAGAGCTCTCAATGAAAGATATGCGGGGTGCTGCAACCCGCTGAGGGAGAGAGCAT 1440  
1441 GCGAACCACTCTCGTGGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
1441 GCGAACCACTCTCGTGGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
1501 CATGGAGAGCCGTGGAGAGAGATGAGAGAGCTTCTCTCAAGAGAGAGAGAGAGAGAGAG 1560  
1501 CATGGAGAGCCGTGGAGAGAGATGAGAGAGCTTCTCTCAAGAGAGAGAGAGAGAGAGAG 1560  
1561 AGGAAAGCCCTCTGAG 1620  
1561 AGGAAAGCCCTCTGAG 1620  
1621 TCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1621 TCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1681 AAGAGCTGAAGAGCTTCTCCCGACCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
1681 AAGAGCTGAAGAGCTTCTCCCGACCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
1741 TGGCCTGTGTGAG 1800  
1741 TGGCCTGTGTGAG 1800  
1801 AATGTTCAAGAGTGTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
1801 AATGTTCAAGAGTGTGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
1861 CCGAGGTGCTTGTGTGAG 1920  
1861 CCGAGGTGCTTGTGTGAG 1920  
1921 GGAACAGATGGAG 1980  
1921 GGAACAGATGGAG 1980  
1981 GCAAGCTCAG 2040  
1981 GCAAGCTCAG 2040  
2041 GACTACAGATTTATTTTAAACAGAGAAATGTTTAAATTTATTAATTCATATTTATA 2100  
2041 GACTACAGATTTATTTTAAACAGAGAAATGTTTAAATTTATTAATTCATATTTATA 2100  
2101 ATGTTGGCCCAACATTAATGATTTCTTGTCTGATCTGATCTGATCTGATCTGATCTGAT 2160  
2101 ATGTTGGCCCAACATTAATGATTTCTTGTCTGATCTGATCTGATCTGATCTGATCTGAT 2160  
2161 GAAAGAACTTTAAACAGATTAATGATGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
2161 GAAAGAACTTTAAACAGATTAATGATGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
2221 AAAAATGTTGTTTAAAGCTTTTAAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2221 AAAAATGTTGTTTAAAGCTTTTAAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2281 GTGAG 2340  
2281 GTGAG 2340  
2341 TTTTAAATTAATTAATTCGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2390  
2341 TTTTAAATTAATTAATTCGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2390













Db	1401	ACCCCTGTAAGGAGCAGAGGCATGTCCACACCACTCCGTGGAGAGTCCAGAGGCCACGAGAGG	1460
Qy	1480	TCCTGGGGGGACCCCATGTGGTCCGATGGGAGGCCGCTGGAGAGATATGAGAGGCTTCCTCC	1539
Db	1461	TCCTGGGGGGACCCCATGTGGTCCGATGGGAGGCCGCTGGAGAGATATGAGAGGCTTCCTCC	1520
Qy	1540	AAGAGGACCCCGGGGGTTCCGAGGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	1599
Db	1511	AAGAGGACCCCGGGGGTTCCGAGGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	1580
Qy	1600	TGGAGACTGGCAGTGTCCCAAATCCGGGATTGGAAAACCAAGACTTCGCTTGGAGAACAG	1659
Db	1581	TGGAGACTGGCAGTGTCCCAAATCCGGGATTGGAAAACCAAGACTTCGCTTGGAGAACAG	1640
Qy	1680	GTGCACCAACAGTGTAAAGGCCCCCAAGCCCTGAAAGGCTTCTCCCCGCACTCTTCCGCCCC	1719
Db	1641	GTGCACCAACAGTGTAAAGGCCCCCAAGCCCTGAAAGGCTTCTCCCCGCACTCTTCCGCCCC	1700
Qy	1720	GGGTGGTATCGTGGCAGAGGGTGGCCCTGTGTGGCATTGGCGGGAGAGAAAGGTGGCTCAT	1779
Db	1701	GGGTGGTATCGTGGCAGAGGGTGGCCCTGTGTGGCATTGGCGGGAGAGAAAGGTGGCTCAT	1760
Qy	1780	GGATCGTGTGGTCCCGGTGGAAATGTTCAAGAGTGGCCGTGTGGAGACAGAGGTGGCTT	1839
Db	1761	GGATCGTGTGGTCCCGGTGGAAATGTTCAAGAGTGGCCGTGTGGAGACAGAGGTGGCTT	1820
Qy	1840	CCGTGTGTGGCCGGGGCAGATGACACCAAGTGTGCTTTGTGTGGAGAGAACAGAGTGGCCCTGG	1899
Db	1821	CCGTGTGTGGCCGGGGCAGATGACACCAAGTGTGCTTTGTGTGGAGAGAACAGAGTGGCCCTGG	1880
Qy	1900	GGGGGCCCTCTGGAACCTTTGATGTGAACAGATGGAGAGAGAAAGAGAGGACCTGGAGGAC	1959
Db	1881	GGGGGCCCTCTGGAACCTTTGATGTGAACAGATGGAGAGAGAAAGAGAGGACCTGGAGGAC	1940
Qy	1960	TGGAAAAATGATAAAGGCGAGCACCGTCAGAGGCGCAGAGATCGGCCCTTACTATGATGCA	2019
Db	1941	TGGAAAAATGATAAAGGCGAGCACCGTCAGAGGCGCAGAGATCGGCCCTTACTATGATGCA	2000
Qy	2020	GAGACCCCGCAGAGCTGATGACATACCAAGATTTATTTTTPAACCCAGAAATGTTTTAA	2079
Db	2001	GAGACCCCGCAGAGCTGATGACATACCAAGATTTATTTTTPAACCCAGAAATGTTTTAA	2060
Qy	2080	ATTATTAATTCATATTATTAATGTTGGCCCAACACATATGATTTATTCCTGTGCTGACT	2139
Db	2061	ATTATTAATTCATATTATTAATGTTGGCCCAACACATATGATTTATTCCTGTGCTGACT	2120
Qy	2140	TTAGTATTTTTCACCATTTGTGAGAGAAACATTTAAACAGATTAAATGTTAGTGTGGGAG	2199
Db	2121	TTAGTATTTTTCACCATTTGTGAGAGAAACATTTAAACAGATTAAATGTTAGTGTGGGAG	2180
Qy	2200	TTTTTTTTTCTCTCTCTTTTAAAAATGTTGTTTAAACATTTAACATGGGAACCCCTT	2259
Db	2181	TTTTTTTTTCTCTCTCTTTTAAAAATGTTGTTTAAACATTTAACATGGGAACCCCTT	2240
Qy	2260	GTGAGCATGCTCAGATCATGTTGTGAGAGAACCAAGAGGCCCTTTAATCTGTAAACATGTT	2319
Db	2241	GTGAGCATGCTCAGATCATGTTGTGAGAGAACCAAGAGGCCCTTTAATCTGTAAACATGTT	2300
Qy	2320	ATGTTGTGATGTTTTTTTTTTTTTTTTTTTTTAAAAATTAATTCCAATGTTTAAATAAAAA	2379
Db	2301	ATGTTGTGATGTTTTTTTTTTTTTTTTTTTTTAAAAATTAATTCCAATGTTTAAATAAAAA	2360
Qy	2380	AAAA 2383	
Db	2361	AAAA 2364	

RESULT 7					
CQ730352					
LOCUS	CQ730352	2326 bp	DNA	linear	PAT 03-FEB-2004
DEFINITION	Sequence	16286	from Patent WO02068579.		
ACCESSION	CQ730352				
VERSION	CQ730352.1	GI:42303963			

KEYWORDS	
SOURCE	
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
JOURNAL	Kits, such as nucleic acid arrays, comprising a majority of humanecons or transcripts, for detecting expression and other uses thereof
FEATURES	Patent: WO 02068579-A 16286 06-SEP-2002; PE Corporation (NY) (US)
source	location/Qualifiers
	1..2326
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
ORIGIN	

Query Match	Similarity	95.4%	Score 2279.2	DB 6	Length 2326
Best Local	Similarity	99.7%	Pred. No. 0		
Match 214	Conservative	0	Mismatches	3	Indels
					Gaps
					3
QY	56	GATTACAGTACCTATAGCCAAAGCTGTCAGCGCAGAGGGCTACAGTGTCTTACACCGCCAG	115		
DB	8	GATTACAGTACCTATAGCCAAAGCTGTCAGCGCAGAGGGCTACAGTGTCTTACACCGCCAG	67		
QY	116	CCCACTCAAGATATGCAACAGACCCAGGCGATATGGGCAACAAAGCTATGGAACTTAT	175		
DB	68	CCCACTCAAGGATATGCAACAGACCCAGGCGATATGGGCAACAAAGCTATGGAACTTAT	127		
QY	176	GGACAGCCCACTGATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGGGAGACC	235		
DB	128	GGACAGCCCACTGATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGGGAGACC	187		
QY	236	GCCATGTGAATCTTCTTATGSAAGGCTCCCACTGTATTACTACTCAATGCGCCCGAG	299		
DB	188	GCCATGTGAATCTTCTTATGSAAGGCTCCCACTGTATTACTACTCAATGCGCCCGAG	247		
QY	296	GCATACAGCCAGCCTGTCCAGGGGTATGGCACTGTGCTTATGATACCAACCACTGCTACA	355		
DB	248	GCATACAGCCAGCCTGTCCAGGGGTATGGCACTGTGCTTATGATACCAACCACTGCTACA	307		
QY	356	GTACACCAACCCAGGCGCTCCATAGCAAGTCAAGTCTGATATGGCACTAGCGTGTAT	415		
DB	308	GTACACCAACCCAGGCGCTCCATAGCAAGTCTGATATGGCACTAGCGTGTAT	367		
QY	416	CCAGCCTATGGGCAAGCAGCAGCAGCAGCACTGCACTCAACAAGCCGAGATGGAAACAAG	475		
DB	368	CCAGCCTATGGGCAAGCAGCAGCAGCAGCACTGCACTCAACAAGCCGAGATGGAAACAAG	427		
QY	476	CCCACTGAGACTAGTCAACTGCAATCTAGCACAGGGGGTTACAAACGAGCCGAGCTTAGGA	535		
DB	428	CCCACTGAGACTAGTCAACTGCAATCTAGCACAGGGGGTTACAAACGAGCCGAGCTTAGGA	487		
QY	536	TATGACACAGATTACTACAGTTATCCCAAGGATACCTGGAGGCTACCCCACTGACCGCACTC	595		
DB	488	TATGACACAGATTACTACAGTTATCCCAAGGATACCTGGAGGCTACCCCACTGACCGCACTC	547		
QY	596	ACTGACCCCACTCTCACTCCCTCTACAGAGCTATTCCTCTACACAGCCGACTAGTTATGAT	655		
DB	548	ACTGACCCCACTCTCACTCCCTCTACAGAGCTATTCCTCTACACAGCCGACTAGTTATGAT	607		
QY	656	CAGACAGTTAATCTCAGCAGAAACCTATGGGCAACGAGCAGCTATGGACAGCAGAGT	715		
DB	608	CAGACAGTTAATCTCAGCAGAAACCTATGGGCAACGAGCAGCTATGGACAGCAGAGT	666		
QY	716	AGCTATGTCAAACAAGCAGCTATGGGCAAGGCTCCCACTATGTTAACCAACCCCAACT	775		
DB	668	AGCTATGTCAAACAAGCAGCTATGGGCAAGGCTCCCACTATGTTAACCAACCCCAACT	727		
QY	776	GGATCTTACAGCCAGCTTCAAGTCAATATACCACAAAGACGAGCTTACGGGCAAGCAG	833		

Db 728 GATCTGATCTAGAGCCACCTGTAGATCCAGATGATATATAGCCAAACAGACGACGACTAGCGGACGACG 787  
 Qy 836 AGTTCAATTCGAGACGACACCCCACTAGACATGGGTGTTTATGGGACGAGAGTCTGGAGGA 895  
 Db 788 AGTTCAATTCGAGACGACACCCCACTAGACATGGGTGTTTATGGGACGAGAGTCTGGAGGA 847  
 Qy 896 TTTTCGAGACGAGAGAGAACCGGAGCATGATGGGCCCTGATTAACCGGGGACAGGGGAGA 955  
 Db 848 TTTTCGAGACGAGAGAGAACCGGAGCATGATGGGCCCTGATTAACCGGGGACAGGGGAGA 907  
 Qy 956 GGGGAGATTGATCGTGGAGCATGACAGAGTGGCGGGAGAGAGACCGGTGGATG 1015  
 Db 908 GGGGAGATTGATCGTGGAGCATGACAGAGTGGCGGGAGAGAGACCGGTGGATG 967  
 Qy 1016 GGCAGGCTGGAGAGGAGGCTGCTCAATTAAGCTGTGGACCCATGATGAAGACCA 1075  
 Db 968 GGCAGGCTGGAGAGGAGGCTGCTCAATTAAGCTGTGGACCCATGATGAAGACCA 1027  
 Qy 1076 GATCTGATCTAGAGCCACCTGTAGATCCAGATGAGAACTCTGACAAACAGTGCATTTAT 1135  
 Db 1028 GATCTGATCTAGAGCCACCTGTAGATCCAGATGAGAACTCTGACAAACAGTGCATTTAT 1087  
 Qy 1136 GTACAGATTAAATGACAGTGTGACTTATGATCTGGCAGACTTCTTTAAGCAGTGT 1195  
 Db 1088 GTACAGATTAAATGACAGTGTGACTTATGATCTGGCAGACTTCTTTAAGCAGTGT 1147  
 Qy 1196 GGGGTTGTTAAGTGAACAAAGAACTGGGCAACCATGATCAATCTAAGCTGGACAG 1255  
 Db 1148 GGGGTTGTTAAGTGAACAAAGAACTGGGCAACCATGATCAATCTAAGCTGGACAG 1207  
 Qy 1256 GAAACGAGAAAGCCAAAGGCGATGCAAGTGTCTATGAAGAACCCACCATGACGACAG 1315  
 Db 1208 GAAACGAGAAAGCCAAAGGCGATGCAAGTGTCTATGAAGAACCCACCATGACGACAG 1267  
 Qy 1316 GCTGCGGTGGAATGTTTATGAGGAAAGATTTTCAAGGAGCAACTTAAAGTCTCCCTT 1375  
 Db 1268 GCTGCGGTGGAATGTTTATGAGGAAAGATTTTCAAGGAGCAACTTAAAGTCTCCCTT 1327  
 Qy 1376 GCTCGGAAGAGCTTCAATGAACAGTATCGGGGTGTCTGCCAACCCCGTGAAGGCGAGA 1435  
 Db 1328 GCTCGGAAGAGCTTCAATGAACAGTATCGGGGTGTCTGCCAACCCCGTGAAGGCGAGA 1387  
 Qy 1436 GGCATGTCACACCACTCCGTGAGGATCCAGAGGCGCCAGAGGCTCTGGGGAGACCCAG 1495  
 Db 1388 GGCATGTCACACCACTCCGTGAGGATCCAGAGGCGCCAGAGGCTCTGGGGAGACCCAG 1447  
 Qy 1496 GGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTTCAAGAGGACCCCGGGGT 1555  
 Db 1448 GGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTTCAAGAGGACCCCGGGGT 1507  
 Qy 1556 TCCCGAGGAAACCTCTCTGAGAGAGAAAGTTCACGACCGAGCTGAGACCTGGCAGTGT 1615  
 Db 1508 TCCCGAGGAAACCTCTCTGAGAGAGAAAGTTCACGACCGAGCTGAGACCTGGCAGTGT 1567  
 Qy 1616 CCCAATCCCGGTTGTGAGAAACCGAACTTCCCTGAGAAACAGAGTGCACACGAGTAAAG 1675  
 Db 1568 CCCAATCCCGGTTGTGAGAAACCGAACTTCCCTGAGAAACAGAGTGCACACGAGTAAAG 1627  
 Qy 1676 GCCCAGAAAGCTGAAAGGCTTCTCCCGGCAACCTTTCGCGCCCGGGGTGTGATCGTGGC 1735  
 Db 1628 GCCCAGAAAGCTGAAAGGCTTCTCTCCGCAACCTTTCGCGCCCGGGGTGTGATCGTGGC 1687  
 Qy 1736 AAGAGTGCCTCTGTGAGCATGCGGGAGAGAAAGGTGGCTCTGATGATCGTGTGCTCC 1795  
 Db 1688 AAGAGTGCCTCTGTGAGCATGCGGGAGAGAAAGGTGGCTCTGATGATCGTGTGCTCC 1747  
 Qy 1796 GGTGGAATGTTGAGAGTGGCGGTGTGAGAGACAGAGTGGCTTCCGTGTGGCCCGGGG 1854  
 Db 1748 GGTGGAATGTTGAGAGTGGCGGTGTGAGAGACAGAGTGGCTTCCGTGTGGCCCGGGG 1807  
 Qy 1855 CATGACCGAGGTGCTTTGTGTGAGAGAAACAGAGTGGCCCTGAGGAGGCCCCCTGAGACC 1914  
 Db 1808 CATGACCGAGGTGCTTTGTGTGAGAGAAACAGAGTGGCCCTGAGGAGGCCCCCTGAGACC 1867

Qy 1915 TTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGAGAGACTTGAAAAATGATTA 1974  
 Db 1868 TTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGAGAGACTTGAAAAATGATTA 1927  
 Qy 1975 AGGCGAGACCCGTCAGAGGCGAGAGATGGGCCCTACTGATGACAGAGACCCCGAGAGC 2034  
 Db 1928 AGGCGAGACCCGTCAGAGGCGAGAGATGGGCCCTACTGATGACAGAGACCCCGAGAGC 1987  
 Qy 2035 TGCATTGACTACAGATTTATTTTAAACAGAAAAATGTTTAAATTAATTCATTA 2094  
 Db 1988 TGCATTGACTACAGATTTATTTTAAACAGAAAAATGTTTAAATTAATTCATTA 2047  
 Qy 2095 TTTATATGTTGGCCCAAC-ATTATGATTTATCTTGTCTGACTTTAGATTTTAC 2153  
 Db 2048 TTTATATGTTGGCCCAACAAATTATGATTAATTCCTGTCTGACTTTAGATTTTAC 2107  
 Qy 2154 CATTTGTAAGAAACATTAATAAGTTAAATGATGAGGAGTTTCTTCCTCC 2213  
 Db 2108 CATTTGTAAGAAACATTAATAAGTTAAATGATGAGGAGTTTCTTCCTCC 2167  
 Qy 2214 TTTCTTTAAAAATGTTGTTTAAAGCTTTAAACATTAAGGAAACCCCTGTGAGCATGCTCAG 2273  
 Db 2168 TTTCTTTAAAAATGTTGTTTAAAGCTTTAAACATTAAGGAAACCCCTGTGAGCATGCTCAG 2227  
 Qy 2274 TATCATTTGGAGAACCAAGAGGCGCTTTAACTGTAACTGAATGTTGATGATGTT 2333  
 Db 2228 TATCATTTGGAGAACCAAGAGGCGCTTTAACTGTAACTGAATGTTGATGATGTT 2287  
 Qy 2334 TTTTCTTTTAAAAATTAATAATTCGAATGTTAAATA 2373  
 Db 2288 TTTTCTTTTAAAAATTAATAATTCGAATGTTAAATA 2326

RESULT 8  
 CR456490  
 LOCUS  
 DEFINITION  
 Homo sapiens EWSR1 full length open reading frame (ORF) cDNA clone  
 (cDNA clone C220RF:pgEM.EWSR1).  
 ACCESSION  
 CR456490  
 VERSION  
 CR456490.1 GI:47678510  
 KEYWORDS  
 CDNA; chromosome 22; ORF.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 Collins, J.E., Wright, C.L., Edwards, C.A., Davis, M.P., Grimham, J.A.,  
 1 (bases 1 to 2189)  
 Cole, C.G., Goward, M.E., Aguado, B., Mallia, M., Mokrad, Y.,  
 Huckle, E.J., Beare, D.M. and Dunham, I.  
 JOURNAL  
 Submitted (24-MAY-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: c220@sanger.ac.uk Manuscript  
 Sanger Institute name : pgEM.EWSR1  
 COMMENT  
 Homo sapiens CDNA sequence. This sequence was generated as part of  
 The Wellcome Trust Sanger Institute program to isolate cDNA clones  
 representing the full length open reading frame of well annotated  
 protein coding genes on human chromosome 22. For more information  
 see <http://www.sanger.ac.uk/HGP/Chr22/>.  
 FEATURES  
 source  
 1..2189  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /clone="pgEM.EWSR1"  
 /lab\_host="JM109"  
 1..2189  
 /gene="EWSR1"  
 34..2001  
 /gene="EWSR1"  
 /codon\_start=1  
 /protein\_id="CAG30376.1"



QY	1931	GGAGGAAAGAGGAGGACGTGGAGGACCTGGAAAAATGGATTAAGGGAGACACCTGAG	1997		
Db	1918	GGAGGAAAGAGGAGGACGTGGAGGACCTGGAAAAATGGATTAAGGGAGACACCTGAG	1977		
QY	1991	GAGCGCAGAGATCGGCGCTTACTAGATGCAGAGACCCCGCAGAGCTGCATTGACTACACGA	2050		
Db	1978	GAGCGCAGAGATCGGCGCTTACTAGATGCAGAGACCCCGCAGAGCTGCATTGACTACACGA	2037		
QY	2051	TTTATTTTAAACCGAAAAATGTTTTAAATTTAAATTCATTTTAAATTTAAATTTGTTGGCA	2110		
Db	2038	TTTATTTTAAACCGAAAAATGTTTTAAATTTAAATTTAAATTTTAAATTTGTTGGCA	2097		
QY	2111	CAACATTATGATTTATTCCTTGCTGTACTTATAGTTATTTTTCACCATTTGTGAAGAACAT	2170		
Db	2098	CAACATTATGATTTATTCCTTGCTGTACTTATAGTTATTTTTCACCATTTGTGAAGAACAT	2157		
QY	2171	TAAACAAAGTTAAATGTAGTGTGGGAGTTT	2202		
Db	2158	TAAACAAAGTTAAATGTAGTGTGGGAGTTT	2189		
RESULT 9	BC004817	2182 bp	mRNA	linear	PRI 29 -JUN-2004
LOCUS	BC004817				
DEFINITION	Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant				
ACCESSION	BC004817				
VERSION	BC004817.1	GI:13435962			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 2182)				
AUTHORS	Strausberg, R. L., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Klausner, R. D., Collins, F. S., Zeeberg, B., Buetow, K. H., Scheef, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Heide, F., Datchenko, L., Marusik, K., Farmer, A. A., Rubin, G. M., Hong, L., Stepieton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Udell, T. B., Yoshinuki, S., Caminci, P., Prange, C. A., Raha, S. S., Loquellano, N. A., Peters, G. J., Adamson, R. D., Mullany, S. J., Bosak, S. A., McEwan, P. J., McEwan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Mowley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hilyk, S. W., Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butlerfield, Y. S., Kizylinski, M. J., Skalski, U., Smalls, D. E., Scherch, A., Schein, J. E., Jones, S. J., and Harris, M. A.				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 2182)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:gsapbse-remail.nih.gov">gsapbse-remail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC				

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnarane, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.U., Martin, R.G., Muzny, D.M., Navavati,  
A.N., Gibbs, R.A.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.lnl.gov>  
Series: IRK Plate: 3 Row: 1 Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

Location/Qualifiers  
1. .2182

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:5432 IMAG:3449145"
/tissue_type="Placenta, choriocarcinoma"
/clone_id="NH MGC_10"
/lab_host="DH10E"
/notes="vector: pCMV-SPORT6"
1..2182
name"
gene

```

```

/note="synonym: EMS"
/db_xref="LocusID:2130"
/db_xref="MIM:133450"
20.1990
CDS

```

**CDS**

```

/product="Bwing sarcoma breakpoint region 1, isoform EMS
/protein_id="AB04817.1"
/db_xref="GI:13435963"
/db_xref="LOCUSID:2130"
/db_xref="MIM:133450"
/translation="MASTDYSTVSSAAAGQGSAYTAQPTQGVAAQTQAYGQGSYGT
GQPTVSYAQQTATATYGTGTAATSYGQPPGTYTTPAPASQPVQVGGAGVADTT
ATVQTAASVAAQSAVGTGPAPVAGQGPATATPRDNDKPKPETSQSSQNTGCG
BLATGQGSYSYPOVPGSYPMQPTTAPTSYPTYSSTQPTSDQSSYPOONTGQ
STGQGSSTGQGSSTGQDPPTSYPTQTSYSAPQSYQGSQSSSTQGSFPDNDHSS
VYGQSSGGSGPGENRSMGSDNRRGRGGRGDMGSGRGGGRGSGAGRGSG
KGGPMDEGPDLDLPVPDPEDSDNSAIYQGLNDSITLDDLPFKQGVYMNNT
TQCPIMHILYDKETGPKGATVSYDEPTKAAVEMFGDKDFGSGLTATSLARKK
NMSMGGLPREGREKMPPLRGGPGGPGGPMGRGGRGGRGCGPFRPGGSGSR
BGGGCVNHRAGDNQCNPGCGGNONFANRTETCNCKAPKBEGLTPPPPPPGGDRG
GFGNRRGGGGLMDRGPGGPMFRGGGRGGGRGGGMDGGGGGRGGRGCGGPP
LNEQGGRRGGGCGKMDKGEHNGQERRDYY"

```

**ORIGIN**

Query Match	90.6%	Score 2165;	DB 9;	Length 2182;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2165; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy	85	GCACGACGGGCTTCACTGCTTTACACCGCCCAAGCCCACTCAGAGATATGACAGACACCCA	144
Db	61	GCACGACGGGCTTCACTATGCTTTACACCGCCCAAGCCCACTCAGAGATATGACAGACACCCA	120
Qy	145	GGCATATTGGGCAACAAAGCTATGGAACTTATGGACAGGCCATGTATGTACAGTATACC	204
Db	121	GGCATATTGGGCAACAAAGCTATGGAACTTATGGACAGGCCATGTATGTACAGTATACC	180
Qy	205	GGCTCAGACCACTGCAACTTATGGCAGACCGCCTTATGCACTTTCTTATGGACAGCCTCC	264
Db	181	GGCTCAGACCACTGCAACTTATGGCAGACCGCCTTATGCACTTTCTTATGGACAGCCTCC	240
Qy	265	CACCTGTTATTAATTCCAACTGCCCCCAGGCAATACAGCAGGCTGTCCAGGGGTTATGG	324
Db	241	CACCTGTTATTAATTCCAACTGCCCCCAGGCAATACAGCAGGCTGTCCAGGGGTTATGG	300

QY 325 CACTGGTGTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTATGACG 384  
 DB 301 CACTGGTGTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTATGACG 360  
 QY 385 TCAGTGTGATATGAGCACTGAGCTGCTTATCCAGCTTATGGGAGCAGCAGCAGC 444  
 DB 361 TCAGTGTGATATGAGCACTGAGCTGCTTATCCAGCTTATGGGAGCAGCAGCAGC 420  
 QY 445 TGGACTACAAAGCCCGAGATGAAAACAGCCCACTGAGCACTAGTCAACCTCAACTG 504  
 DB 421 TGGACTACAAAGCCCGAGATGAAAACAGCCCACTGAGCACTAGTCAACCTCAACTG 480  
 QY 505 CACAGGAGGTTTAAACACAGCCAGCTAGATATGACAGAGTAACTACAGTTATCCCA 564  
 DB 481 CACAGGAGGTTTAAACACAGCCAGCTAGATATGACAGAGTAACTACAGTTATCCCA 540  
 QY 565 GGTACTGGAGGTATCCCACTGAGCAGCAGTCACTGCACTCCATCCCTCTACAG 624  
 DB 541 GGTACTGGAGGTATCCCACTGAGCAGCAGTCACTGCACTCCATCCCTCTACAG 600  
 QY 625 CTATTCCTCTTACACAGCCCACTAGTATGATCAAGAGTAACTCTCAGAGAAACCTTA 684  
 DB 601 CTATTCCTCTTACACAGCCCACTAGTATGATCAAGAGTAACTCTCAGAGAAACCTTA 660  
 QY 685 TGGGCAACCGAGAGCTATGAGCAGAGTAACTATGTCACAAAGCAGTATGAGCA 744  
 DB 661 TGGGCAACCGAGAGCTATGAGCAGAGTAACTATGTCACAAAGCAGTATGAGCA 720  
 QY 745 GCAAGCTCCCACTAGTTATCCCACTGAGCAGCAGTCACTCCATCCCTCTACAG 804  
 DB 721 GCAAGCTCCCACTAGTTATCCCACTGAGCAGCAGTCACTCCATCCCTCTACAG 780  
 QY 805 TACCCAAACAGAGAGCTACCGGCGAGCAGAGTCACTCCCAAGAGCAACCCAGTGG 864  
 DB 781 TACCCAAACAGAGAGCTACCGGCGAGCAGAGTCACTCCCAAGAGCAACCCAGTGG 840  
 QY 865 CATGGGTGTTTATGGGCGAGAGTCTGAGAGATTTTCCGACACAGAGAAACCGGAGAT 924  
 DB 841 CATGGGTGTTTATGGGCGAGAGTCTGAGAGATTTTCCGACACAGAGAAACCGGAGAT 900  
 QY 925 GAGTGGCTCTGATTAACCGGCGAGAGGAGGAGATTTGATCTGAGAGCAGTGAAG 984  
 DB 901 GAGTGGCTCTGATTAACCGGCGAGAGGAGGAGATTTGATCTGAGAGCAGTGAAG 960  
 QY 985 AGGTGGCGGGGAGAGAGAGCGGCTGAGATGGGCGAGCAGAGAGGAGGAGTTCGA 1044  
 DB 961 AGGTGGCGGGGAGAGAGAGCGGCTGAGATGGGCGAGCAGAGAGGAGGAGTTCGA 1020  
 QY 1045 TAAAGCTGTGAGCCCATGATGAGAGCAGATCTTGAATCTAGGCCCATCTGATGCC 1104  
 DB 1021 TAAAGCTGTGAGCCCATGATGAGAGCAGATCTTGAATCTAGGCCCATCTGATGCC 1080  
 QY 1105 AGATGAAGACTCTGACAAACAGTGAATTTATGTCAGAGATTAATGACAGTGTACTCT 1164  
 DB 1081 AGATGAAGACTCTGACAAACAGTGAATTTATGTCAGAGATTAATGACAGTGTACTCT 1140  
 QY 1165 AGATGAAGACTCTGACAAACAGTGAATTTATGTCAGAGATTTAAGATGAACAGAGACTGG 1224  
 DB 1141 AGATGAAGACTCTGACAAACAGTGAATTTATGTCAGAGATTTAAGATGAACAGAGACTGG 1200  
 QY 1225 GCAACCATGATCACTATCTGAGCAAGAGAAACAGAGAAAGCCCAAGGCGATGCCAC 1284  
 DB 1201 GCAACCATGATCACTATCTGAGCAAGAGAAACAGAGAAAGCCCAAGGCGATGCCAC 1260  
 QY 1285 AGTGTCTTATGAGAGCCCACTGCAAGGCTGCGGTGAGATGTTTATGAGGAGAAAGA 1344  
 DB 1261 AGTGTCTTATGAGAGCCCACTGCAAGGCTGCGGTGAGATGTTTATGAGGAGAAAGA 1320  
 QY 1345 TTTTCAAGGAGAGCAAAATTTAAAGTCTCCCTTGCCTGGAGAAAGGCTCCCAATGAACGATAT 1404  
 DB 1321 TTTTCAAGGAGAGCAAAATTTAAAGTCTCCCTTGCCTGGAGAAAGGCTCCCAATGAACGATAT 1380  
 QY 1405 GCGGGGTGTCTGCAACCCGCTGAGGCGAGAGGATGCCACCACTCCGTGGAGGTTC 1464

DB 1381 GCGGGGTGTCTGCAACCCGCTGAGGCGAGAGGATGCCACCACTCCGTGAGGTCC 1440  
 QY 1465 AGGAGGCCCAAGAGGCTCTGGGGGAGACCAATGGGTGGATGGAGGCCGTGGAGAGATG 1524  
 DB 1441 AGGAGGCCCAAGAGGCTCTGGGGGAGACCAATGGGTGGATGGAGGCCGTGGAGAGATG 1500  
 QY 1525 AGGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAA 1584  
 DB 1501 AGGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAA 1560  
 QY 1585 CGTCCAGACCCGAGCTGAGAGCTGCGAGTGTCCCAATCCGGTGTGGAAACCAACTT 1644  
 DB 1561 CGTCCAGACCCGAGCTGAGAGCTGCGAGTGTCCCAATCCGGTGTGGAAACCAACTT 1620  
 QY 1645 CGCCTGGAGAAACAGAGTGCACCACTGTAAAGGCCCAAGGCTTCCCTCCGCGC 1704  
 DB 1621 CGCCTGGAGAAACAGAGTGCACCACTGTAAAGGCCCAAGGCTTCCCTCCGCGC 1680  
 QY 1705 ACCCTTCCGCCCCCGGGGTGATCGTGGCAGAGGTGACCCTGTGCGATGCGGGAGG 1764  
 DB 1681 ACCCTTCCGCCCCCGGGGTGATCGTGGCAGAGGTGACCCTGTGCGATGCGGGAGG 1740  
 QY 1765 AAGAGGTGCTCATGATCGTGTGTGCTCCGCTGGAATGTTCAAGTGTGCGCTGTGG 1824  
 DB 1741 AAGAGGTGCTCATGATCGTGTGTGCTCCGCTGGAATGTTCAAGTGTGCGCTGTGG 1800  
 QY 1825 AAGCAGAGTGTGCTTCCGTGTGTGCGCGGGAGCAGACCGAGGTGCTTGTGGAGAGAG 1884  
 DB 1801 AAGCAGAGTGTGCTTCCGTGTGTGCGCGGGAGCAGACCGAGGTGCTTGTGGAGAGAG 1860  
 QY 1885 ACGAGGTGCTCCGCGGGGAGCCTTGAGACCTTGTGATGAAACAGATGGAGAGAGAG 1944  
 DB 1861 ACGAGGTGCTCCGCGGGGAGCCTTGAGACCTTGTGATGAAACAGATGGAGAGAGAG 1920  
 QY 1945 AAGACGTGAGAGACCTGGAATAATGATTAAGGCGAGACCCGTCAAGAGCGCAGAGATGG 2004  
 DB 1921 AAGACGTGAGAGACCTGGAATAATGATTAAGGCGAGACCCGTCAAGAGCGCAGAGATGG 1980  
 QY 2005 GCCCTATGATGAGAGAGACCCCGAGAGCTCATGATGATCAAGATTTATTTTAAAC 2064  
 DB 1981 GCCCTATGATGAGAGAGACCCCGAGAGCTCATGATGATCAAGATTTATTTTAAAC 2040  
 QY 2065 CAGAAATGTTTAAATTTATTAATTCATATTTATATGTTGGCCCAACATTAATGATTA 2124  
 DB 2041 CAGAAATGTTTAAATTTATTAATTCATATTTATATTAATTTATTAATTTATTAAT 2100  
 QY 2125 TTCTTGTCTGATCTTAAATTTTACCATTTTGTGAAGAAACATTAACCAAGTTTAA 2184  
 DB 2101 TTCTTGTCTGATCTTAAATTTTACCATTTTGTGAAGAAACATTAACCAAGTTTAA 2160  
 QY 2185 TGGTA 2189  
 DB 2161 TGGTA 2165

RESULT 10  
 BC072442 2164 bp mRNA linear PRI 30-JUN-2004  
 LOCUS Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 DEFINITION B072442  
 ACCESSION BC072442  
 VERSION BC072442.1 GI:48734726  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2164)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Dege, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Wang, J., Wang, J., Heide, F.,













Db 30624 GGGAGAGCCAGTGAAGCACTGCACTTACAGAACCGGAGAAATGGAACACACCCCACTGAGCA 30565  
 Qy 486 CTAGTCAACTCAATCTAGACAGAGGGGTTTAAACACGCCAGCCTTAGATATGACAGA 545  
 Db 30564 CTAGTCAACTCAATCTAGACAGAGGGGTTTAAACACGCCAGCCTTAGATATGACAGA 30505  
 Qy 546 GTTACTCAAGTTATCTCCAGAGTACCTGGAGGCTAACCCCATATGCAAGCCATGACCTC 605  
 Db 30504 GTTACTCAAGTTATCTCCAGAGTACCTGGAGGCTAACCCCATATGCAAGCCATGACCTC 30445  
 Qy 606 CATCTACCTCTCTACAGCTATTCCTTACACAGCCGAGCTAGTTATGATCAGAGAGATT 665  
 Db 30444 CATCTACCTCTCTACAGCTATTCCTTACACAGCCGAGCTAGTTATGATCAGAGAGATT 30385  
 Qy 666 ACTCTAGACAGAACCTTATGGGCAACCGAGAGCTATGAGACAGAGAGTATGTC 725  
 Db 30384 ACTCTAGACAGAACCTTATGGGCAACCGAGAGCTATGAGACAGAGAGTATGTC 30325  
 Qy 726 AACAAAGCAGCTATGGGAGAGACCTTCCCTAGTTAACCAACCCCAATGATCCTTACA 785  
 Db 30324 AACAAAGCAGCTATGGGAGAGACCTTCCCTAGTTAACCAACCCCAATGATCCTTACA 30265  
 Qy 786 GCCAAGCTCCAACTGCAATATAGCCAAAGAGAGAGCTACGGGCAAGAGATTCAATCC 845  
 Db 30264 GCCAAGCTCCAACTGCAATATAGCTTAAAGAGAGAGCTACGGGCAAGAGATTCAATCC 30205  
 Qy 846 GACAGAGACCAACCCAGTATGATGGGTGTTTATGGGAGAGAGTCTGAGAGATTTCGGAC 905  
 Db 30204 GACAGAGACCAACCCAGTATGATGGGTGTTTATGGGAGAGAGTCTGAGAGATTTCGGAC 30145  
 Qy 906 CAGAGAGAACCCGAGATGATGGGCTCTGATTAACCGGGGAGAGGAGAGGGGATTG 965  
 Db 30144 TAGGAGAGAACCCGAGATGATGGGCTCTGATTAACCGGGGAGAGGAGAGGGGATTG 30085  
 Qy 966 ATCTGAGAGAGATGAGAGAGTGGGCGGGGAGAGAGAGCGGGTGAATGGGAGAGGCTG 1025  
 Db 30084 ATCTGAGAGAGATGAGAGAGTGGGCGGGGAGAGAGAGCGGGTGAATGGGAGAGGCTG 30025  
 Qy 1026 GAGAGAGAGTGGCTTCAATTAAGCTGTGAGACCAATGATGAAGAGACAGATCTTGATC 1085  
 Db 30024 GAGAGAGAGTGGCTTCAATTAAGCTGTGAGACCAATGATGAAGAGACAGATCTTGATC 29965  
 Qy 1086 TAGGCCCACTGTAGATCCAGATGAAGATCTTCAACACAGTGCATTTATGACAGAGAT 1145  
 Db 29964 TAGGCCCACTGTAGATCCAGATGAAGATCTTCAACACAGTGCATTTATGACAGAGAT 29905  
 Qy 1146 TAAATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTTAAGCAGTGGGGTGTGTA 1205  
 Db 29904 TAAATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTTAAGCAGTGGGGTGTGTA 29845  
 Qy 1206 AGATGAACAAGAGACTGGGCAACCCATGATCCATCTTACCTGACAAAGAAACAGAA 1265  
 Db 29844 AGATGAACAAGAGACTGGGCAACCCATGATCCATCTTACCTGACAAAGAAACAGAA 29785  
 Qy 1266 AGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCAACCCACTGCCAAGGCTGGCTGG 1325  
 Db 29784 AGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCAACCCACTGCCAAGGCTGGCTGG 29725  
 Qy 1326 AATGTTTATGAGGAAAGATTTTCAAGGGGAGCAACTTAAAGCTCCCTTGCTCGGAGAGA 1385  
 Db 29724 AATGTTTATGAGGAAAGATTTTCAAGGGGAGCAACTTAAAGCTCTCTTGCTCGGAGAGA 29665  
 Qy 1386 AGCCTTCAATGAACACTATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGGAGATCCAC 1445  
 Db 29664 AGCCTTCAATGAACACTATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGGAGATCCAC 29605  
 Qy 1446 CACCACTCCGTGAGAGTCCAGAGAGGCCAGAGAGTCTTGGGGAGCCCATAGGCTGCATGG 1505  
 Db 29604 CACCACTCCGTGAGAGTCCAGAGAGGCCAGAGAGTCTTGGGGAGCCCATAGGCTGCATGG 29545  
 Qy 1506 GAGGCGGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCGGGGGTTCCTCGAGAGGA 1565

Db 29544 GAGGCCGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCAGGTTCCCGAGGGA 29485  
 Qy 1566 ACCCTCTGAGAGAGGAAAAGTTCAGACACGAGCTGAGAGACTGGATGCCATCCGG 1625  
 Db 29484 ACACTCTGAGAGAGGAAAAGTTCAGACACGAGCTGAGAGAGGAGAGTGTCCCAATCCGG 29425  
 Qy 1626 GTTGTGAAAACCGAACTTGTGCTGAGAGAACAGATGCAACCATGTATAGGCCCAAGC 1685  
 Db 29424 GTTGTGAAAACCGAACTTGTGCTGAGAGAACAGATGCAACCATGTATAGGCCCAAGC 29365  
 Qy 1686 CTGAAGGCTTCTCCCGCAACCTTCCGGCCCCGGGGTGTATCGGAGAGAGTGGCC 1745  
 Db 29364 CTGAAGGCTTCTCCCGCAACCTTCCCGCCCCGGGGTGTATCGGAGAGAGTGGCC 29305  
 Qy 1746 CTGGTGGCATGCGGGAGAGAGAGTGGCTCTCATGATCGTGTGTCCCGTGAATGT 1805  
 Db 29304 CTGGTGGCATGCGGGAGAGAGAGTGGCTCTCATGATCATGTGTGTCCCGGTGAATGT 29245  
 Qy 1806 TCGAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGATGACCCGAG 1865  
 Db 29244 TCGAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGG-CTGGGATGACCCGAG 29186  
 Qy 1866 GTGGCTTTGTGGAGAGACGAGTGGCCCTGGGGGGCCCTTGACCTTGATGAGAAC 1925  
 Db 29186 GTGGCTTTGTGGAGAGACGAGTGGCCCTGGGGGGCCCTTGACCTTGATGATTAAC 29127  
 Qy 1926 AGATGGAGAGAGAGAGAGAGAGCTGAGAGACTTGGAAAAATGATTAAGGCGAGACCC 1985  
 Db 29126 CATGGAGAGAGAGAGAGAGAGAGCTGAGAGACTTGGAAAAATGATTAAGGCGAGACCT 29067  
 Qy 1986 GTGAGAGCGGACAGAGATCGGCCCTTACTAGATGACAGACCCCGAGAGCTGATGACTA 2045  
 Db 29066 GTGAGAGCGGACAGAGATCGGCCCTTACTAGATGACAGAACCCGACAGCTGATGACTA 29007  
 Qy 2046 CCAGATTTATTTTAAACAGAAATGTTTAAATTAATTCATATTATTAATGTT 2105  
 Db 29006 CCAGATTTATTTTAAACAGAAATGTTTAAATTAATTCATATTATTAATGTT 28947  
 Qy 2106 GGCACACATTAATGATTAATCTTGTCTGTACTTATGATTTTTCACCATTTGTGAAG 2165  
 Db 28946 GGCACACATTAATGATTAATCTTGTCTGTACTTATGATTTTTCACCATTTGTGAAG 28887  
 Qy 2166 AACATTTAAACAGTTAAATGCTA 2189  
 Db 28886 AACATTTAAACAGTTAAATGCTA 28863

RESULT 13  
 AC018774 155815 bp DNA linear HTG 16-MAR-2000  
 LOCUS AC018774  
 DEFINITION Homo sapiens clone RP11-2003, WORKING DRAFT SEQUENCE, 3 unordered  
 pieces.  
 AC018774  
 AC018774.3 GI:7107978  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 155815)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 JOURNAL Homo sapiens, clone RP11-2003  
 TITLE Unpublished  
 REFERENCE 2 (bases 1 to 155815)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Boguslavsky, L., Boucknight, B., Brown, A., Burkett, G., Casle, A.,  
 Chespe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,  
 Dearrellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K.,

6 GAGACGACGTTGAGAGAACGAGGAGGAAGAGAGAAATGGCGTCCACGGATTACAGTA 65

Db	10349	GATGGGGACCTTGAAGAAACGAGAGAAAGAGAGAAATGGCATTCATGATTACAGTA	10408
Qy	66	CTTATAGCCAAAGCTGCAAGCGGACGACAGGGCTCAAGTGGCTTAAACCGGCCAGCCACTCAAG	125
Db	10409	CTTATAGCCAAAGCTGCAAGCGGACGACAGGGGGTTACAGTGGCTTAAACCGGCCAGCTCAAG	10468
Qy	126	GATATGACAGACCAACCCAGGGCACTATGGGCAACAAAGCTATGGAACCTATGACAGGCCA	185
Db	10469	GATATGACAGACCAACCCAGGGCACTATGGGCAACAAAGCTATGGAACCTATGACAGGCCA	10528
Qy	186	CTGATGTCAAGCTTATACCCAGGCTCAGAACCACTGCACCTATGGGAGACCGGCTTATGCA	245
Db	10529	TTGATGTCAAGCTTATACCCAGGCTCAGAACCACTGCACCTATGGGAGACCGGCTTATGCA	10588
Qy	246	CTTCTTATGGAACAGCCTCCCACTGGTTATACTCTCAACTGGCCCCCGGCACTACAGCC	305
Db	10589	CTTCTTATGGAACAGCCTCCCACTGGTTATACTCTCAACTGGCCCCCGGCACTACAGCC	10648
Qy	306	AGCGTGCACGGGGTATGCACTGGGGCTTATGATACCACTGCACCTGCTACAGTACCAACA	365
Db	10649	AGCGTGCACGGGGTATGCACTGGGGCTTATGATACCACTGCACCTGCTACAGTACCAACA	10708
Qy	366	CCGAGGCTCTCTATGAGAGCTCAGTCTGCATATGCACTCAGGCTGCTTATCCAGGCTATG	425
Db	10709	CCGAGGCTCTCTATGAGAGCTCAGTCTGCATATGCACTCAGGCTGCTTATCCAGGCTATG	10768
Qy	426	GGCAGCAGCCAGACCACTGCACCTACCAAGACCGGAGATGGAACAAAGCCCACTGACA	485
Db	10769	GGCAGCAGCCAGATGACCACTGCACCTACCAAGACTGCAGATGGAACAAAGCCCACTGAGA	10828
Qy	486	CTTATGAACTCTCAATCTAGACAGGGGGGTAAACAAGGCCAGGCTGAGATATGACAGA	545
Db	10829	CTTATGAACTCTCAATCTAGACAGGGGGGTAAACAAGGCCAGGCTGAGATATGACAGA	10888
Qy	546	GTAATCTACAGTTATCCCGAGGTACTGGGAGCTACCCCATGAGCCACTGCATGCACCTC	605
Db	10889	GTAATCTACAGTTATCCCGAGGTACTGGGAGCTACTCCATGACGCAAGTCAACGACCTC	10948
Qy	606	CATCTTACCTCTCAACAGCTATTTCTTTCATCACAGCCGACTAGTTATGATCAGAGCATTT	665
Db	10949	CATCTTACCTCTCAACAGCTATTTCTTTCATCACAGCCCACTAGTTATGATCAGAGCATTT	11008
Qy	666	ACTCTGACGAAACACCTATGGGCAACGAGCAGCTATGAGCAGCAGATAGCTATGGTC	725
Db	11009	ACTCTGACGAAACACCTATGGGCAACGAGCAGCTATGAGCAGCAGATAGCTATGGTC	11068
Qy	726	AACAAACAGCTATGGGACGACCTTCCACTAGTTAACCAACCCCAATCTGATCCTACA	785
Db	11069	AACAAACAGCTATGGGACGACCTTCCACTAGTTAACCAACCCCAATCTGATCCTACA	11128
Qy	786	GCCAAAGTCCAAGTCAATATATAGCCAAACAGACGACGACTACGGGACGACAGATCTATTC	845
Db	11129	GCCAAAGTCCAAGTCAATATATAGCCAAACAGACGACGACTACGGGACGACAGATCTATTC	11188
Qy	846	GACAGACCAACCCAGTATAGATGGGTGTTATATGGGAGAGATCTGAGGATTTTCCGGAC	905
Db	11189	GACAGACCAACCCAGTATAGATGGGTGTTATATGGGAGAGATCTGAGGATTTTCCGGAC	11248
Qy	906	CAGAGAGAAACCGGAGCATGATGTGCGCTCTGATACCGGGGACAGGGGAAAGGGGGGATTTG	965
Db	11249	TAGAGAGAAACCGGAGCATGATGTGCGCTCTGATATCTGGGGACAGGGGAAAGGGGGATTTG	11308
Qy	966	ATCTGTGAGGACATGACGAGAGTGGGCGGGGAGAGAGACGCGGTGGAATGGGACGCGCTG	1025
Db	11309	ATCTGTGAGGAGATGACGAGAGTGGGCGGGGAGAGAGATGCGGTGGAATGGGACGCGCTG	11368
Qy	1026	GAGAGCGAGGTGGCTTCAATTAAGCTGTGTGAGCCCATGATGAAAGAACAGATCTTGAATC	1085
Db	11369	GAGAGCGAGGTGGCTTCAATTAAGCTGTGTGAGCCCATGATGAAAGAACAGATCTTGAATC	11428
Qy	1086	TAGGCCCACTGTATGATTCAGATGAAAGACTCTGACAAACGTGCAATTTTATGTCAAGGAT	1145
Db	11429	TAGGCCCACTGTATGATTCAGATGAAAGACTCTGACAAACGTGCAATTTTATGTCAAGGAT	11488

QY 1146 TAAATGACAGTGTGACTCTAGATGATGCTGGCAGACTTCTTTAAAGCAGTGGGGTTCTTA 1205  
 DB 11489 TAAATGACAGTGTGACTCTAGATGATGCTGGCAGACTTCTTTAAAGCAGTGGGGTTCTTA 11548  
 QY 1206 AGATGAACAAGAGAACTGGGCAACCCATGATCCATCTCTGAGCAAGAGAAACAGGAA 1265  
 DB 11549 AGATGAACAAGAGAACTGGGCAACCCATGATCCATCTCTGAGCAAGAGAAACAGGAA 11608  
 QY 1266 AGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGACCCAGGCTGCCGTGG 1325  
 DB 11609 AGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGACCCAGGCTGCCGTGG 11668  
 QY 1326 AATGTTTGAATGGAAAAGATTTCAAGGGAGCAAACTTAAGTCTCCCTGCTCGGAAGA 1385  
 DB 11669 AATGTTTGAATGGAAAAGATTTCAAGGGAGCAAACTTAAGTCTCCCTGCTCGGAAGA 11728  
 QY 1386 AGCCTCAATGAACAGTATCGGGGTGTCTGCAACCCCGTGAAGGGGAGAGGATGCGCAC 1445  
 DB 11729 GGCCTCAATGAACAGTATCGGGGTGTCTGCAACCCCGTGAAGGGGAGAGGATGCGCAC 11788  
 QY 1446 CACCACTCCGTGAAGTCCAGAGGCCCAAGAGGTCTGGGGAGACCCATGGGTGCGATGG 1505  
 DB 11789 CACCACTCCGTGAAGTCCAGAGGCCCAAGAGGTCTGGGGAGACCCATGGGTGCGATGG 11848  
 QY 1506 GAGGCGGTGAAGAGATAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTTCGCCAGGGA 1565  
 DB 11849 GAGGCGGTGAAGAGATAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTTCGCCAGGGA 11908  
 QY 1566 ACCCCTCTGAGAGAGAGAAAGTCTGAGACCCGAGCTGAGACTGGAGTGTCCCAATCCGG 1625  
 DB 11909 ACCCCTCTGAGAGAGAGAAAGTCTGAGACCCGAGCTGAGACTGGAGTGTCCCAATCCGG 11968  
 QY 1626 GTTGTGAAAACAGAACTTCTGCTGAGAAACAGAGTGCAACAGTGTAAAGCCCAAGC 1685  
 DB 11969 GTTGTGAAAACAGAACTTCTGCTGAGAAACAGAGTGCAACAGTGTAAAGCCCAAGC 12028  
 QY 1686 CTGAAGAGCTTCCCTCCGCGCAACCTTCCGCGCGGTGTGATCTGTGAGAGAGTGGCC 1745  
 DB 12029 CTGAAGAGCTTCCCTCCGCGCAACCTTCCGCGCGGTGTGATCTGTGAGAGAGTGGCC 12088  
 QY 1746 CTGTGTGAGAGTCCGGGAGAGAGAGTGGCTCATGTAGTGTGTGCTCCGGGTGAATGT 1805  
 DB 12089 CTGTGTGAGAGTCCGGGAGAGAGAGTGGCTCATGTAGTGTGTGCTCCGGGTGAATGT 12148  
 QY 1806 TCAAGAGTGGCCCTGTGTGAGAGACAGAGTGGCTTCCGTGTGAGTCCGGGAGTGAAC 1865  
 DB 12149 TCAAGAGTGGCCCTGTGTGAGAGACAGAGTGGCTTCCGTGTGAGTCCGGGAGTGAAC 12207  
 QY 1866 GTGGCTTTGTGTGAGAGAGACAGAGTGGCTTCCGTGTGAGTCCGGGAGTGAAC 1925  
 DB 12208 GTGGCTTTGTGTGAGAGAGACAGAGTGGCTTCCGTGTGAGTCCGGGAGTGAAC 12266  
 QY 1926 AGATGGAAGAAAGAGAGAGAGAGTGGAGAGCTGAGAAAGATGATTAAGGAGAGACAC 1985  
 DB 12267 AGATGGAAGAAAGAGAGAGAGAGTGGAGAGCTGAGAAAGATGATTAAGGAGAGACAC 12326  
 QY 1986 GTCAAGAGGCGAGAGATGAGGCTCTTATGATGAGAGACCCGAGAGGCTGCAATGACTA 2045  
 DB 12327 GTCAAGAGGCGAGAGATGAGGCTCTTATGATGAGAGACCCGAGAGGCTGCAATGACTA 12386  
 QY 2046 CGAGATTTATTTTAAACCAAGAAATGTTTTAAATTTATTAATTCATTTATTAATGTT 2105  
 DB 12387 CGAGATTTATTTTAAACCAAGAAATGTTTTAAATTTATTAATTCATTTATTAATGTT 12446  
 QY 2106 GGCACCAATATATGATATCTTGTCTGTACTTATGATTTTCAACATTTTGAAGA 2165  
 DB 12447 GGCACCAATATATGATATCTTGTCTGTACTTATGATTTTCAACATTTTGAAGA 12506  
 QY 2166 AACATTTAAACAGATTAAATGTTA 2189  
 DB 12507 AACATTTAAACAGATTAAATGTTA 12530

RESULT 14  
 AL596087/c  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP11-9L18 on chromosome 1, complete  
 sequence.  
 ACCESSION  
 AL596087  
 VERSION  
 AL596087.11 GI:19351911  
 KEYWORDS  
 HTG.  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (10-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk  
 On Mar 11, 2002 this sequence version replaced gi:1694185.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; Sw:  
 SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr1  
 RP11-9L18 is from the library RPCI-11.1 constructed by the group of  
 Pieker de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6  
 This sequence is the entire insert of clone RP11-9L18 The true  
 end of clone RP11-478J7 is at 136234 in this sequence. The true  
 right end of clone RP11-375H19 is at 73742 in this sequence.  
 Location/Qualifiers  
 1..180718  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-9L18"  
 /clone\_1lb="RPCI-11.1"  
 23201..23229  
 /note="Sequence confirmed by AC011221 and AC018774  
 sequenced by WIR."

misc\_feature  
 complement(23230..23297)  
 /note="Unsure [x] Misc feature Then select the text  
 for the note(s) - [ ] Random repeat [ ] Single clone  
 region [ ] Forced join [x] Other Add a comment here -  
 Sequence from AC011221 and AC018774 sequenced by WIR."  
 complement(46397..46442)  
 /note="Single clone region. Sequence from reads from a  
 short insert library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."  
 complement(102116..102124)  
 /note="767 bases of ISI transposon (X52534) removed here.  
 This sequence represents the duplicated flanking sequence  
 of the ISI."



OY	2106	GCGCACAACATTAAGTATTTCCTGTCGTACTTAGATTTTCAACCATTGTGAAGA	2165
DB	155352	GCCCAACAATTATGATATATCTTCCTCTGCTACTTAGATTTTTTCAACCATTGTGAAGA	1554737
OY	2166	MACATTAAACAAGTTAATGATA	2189
DB	155472	MACATTAAACAAGTTAATGATA	155449
RESULT 15			
ACOL1221			
LOCUS			
DEFINITION	ACOL12221	182501 bp DNA 1linear HTG 04-SEP-2000	
		Homo sapiens clone RP11-9L18, WORKING DRAFT SEQUENCE, 13 unordered	
pieces.			
ACOL1221			
VERSION	ACOL1221.5	GI:9966283	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1	(bases 1 to 182501)	
AUTHORS	Birren,B., Linton,L., Nussbaum,C. and Lander,E.		
TITLE	Homo sapiens, clone RP11-9L18		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 182501)	
AUTHORS	Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,		
	Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Boughalter,B.,		
	Brown,A., Casale,A., Colangelo,M., Collins,S., Collymore,A.,		
	Cooke,P., Dearrellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,		
	Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,		
	Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,		
	Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,T.,		
	Lehoczky,T., Lien,C., Locke,K., MacDonald,P., Marquis,N.,		
	McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrim,J.,		
	Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,		
	Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,		
	Stearns-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,		
	Testaye-S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,		
	Wymen,D., Ye,W.J., Zimmer,A. and Zody.M.		
COMMENT	Direct Submission		
	Submitted (03-OCT-1999) Whitehead Institute/MIT Center for Genome		
	Research, 320 Charles Street, Cambridge, MA 02141, USA		
	On Sep 4, 2000 this sequence version replaced gi:7341864.		
	All repeats were identified using RepeatMasker:		
	Smit, A.P.A. & Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	----- Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIBR		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence_submissions@genome.wi.mit.edu		
	----- Project Information		
	Center project name: LJ016		
	Center Clone name: 9_L_18		
	----- Summary Statistics		
	Sequencing vector: M13; M77815; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Assembly program: Phrap; version 0.960731		
	Consensus quality: 175863 bases at least Q40		
	Consensus quality: 179186 bases at least Q30		
	Consensus quality: 180361 bases at least Q20		
	Insert size: 182000; agarose-fp		
	Insert size: 181301; sum-of-contigs		
	Quality coverage: 5.1 in Q20 bases; agarose-fp		
	Quality coverage: 5.1 in Q20 bases; sum-of-contigs		
	----- NOTE: This is a 'working draft' sequence. It currently		
	* consists of 13 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		

```

* as soon as it is available and the accession number will
* be preserved.
1
6942: contig of 6941 bp in length
*
6942: gap of 100 bp
*
7042: contig of 1061 bp in length
*
8103: gap of 100 bp
*
8203: contig of 1404 bp in length
*
9607: gap of 100 bp
*
9707: contig of 3537 bp in length
*
13244: gap of 100 bp
*
13343: contig of 5812 bp in length
*
19156: gap of 100 bp
*
19256: contig of 8795 bp in length
*
28051: gap of 100 bp
*
28151: gap of 100 bp
*
28151: contig of 11334 bp in length
*
39485: gap of 100 bp
*
39485: contig of 15271 bp in length
*
54856: gap of 100 bp
*
54956: gap of 100 bp
*
54956: contig of 13273 bp in length
*
68229: gap of 100 bp
*
68329: gap of 100 bp
*
68329: contig of 20214 bp in length
*
88543: gap of 100 bp
*
88543: contig of 20700 bp in length
*
88643: gap of 100 bp
*
109343: gap of 100 bp
*
109343: contig of 26676 bp in length
*
136119: gap of 100 bp
*
136219: contig of 46283 bp in length.
FEATURES
SOURCE
1..182501
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-9L1A8"
/clone_1fb="RPC1-11 Human Male BAC"
1..6951
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
7042..8102
/note="assembly_fragment"
8203..9606
/note="assembly_fragment"
9707..13243
/note="assembly_fragment"
13344..19155
/note="assembly_fragment"
19256..28050
/note="assembly_fragment"
28151..39484
/note="assembly_fragment"
39585..54855
/note="assembly_fragment"
54956..68228
/note="assembly_fragment"
68329..88542
/note="assembly_fragment"
88643..109342
/note="assembly_fragment"
109443..136118
/note="assembly_fragment"
136219..182501
/note="assembly_fragment"
clone_end:T7
vector_side:right"
ORIGIN
Query Match 85.1%; Score 203.6; DB 2; Length 182501;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 2103; Conservative 0; Mismatches 79; Indels 2; Gaps 2;
GAGACGACGCTTGAAGAAACGACGAGGAAAGAGAGAAATGCGTCCACGAGTTACAGTA 65
GATGGGACCTTGAAGAAACGAGAGGAAAGAGAGAAATGCGATCCATGAGTTACAGTA 56503

```



QY 66 CCTATAGCCAAAGCTGAGGCGAGGCGGTAGAGTCTTACACCGCCAGCCCACTCAAG 125  
DB 56504 CCTATAGCCAAAGCTGAGGCGAGGCGGTAGAGTCTTACACCGCCAGCCCACTCAAG 56563  
QY 126 GATATGACAGACCAAGGAGATATGAGCAAAAGCTATGGAACCTATGACAGCCCA 185  
DB 56564 GATATGACAGACCAAGGAGATATGAGCAAAAGCTATGGAACCTATGACAGCCCA 56623  
QY 186 CTGATGTCAGTATACCAAGGCTCAGACCACTGCAACCTATGAGGAGACCGCTTATGCA 245  
DB 56624 TTGATGTCAGTATACCAAGGCTCAGACCACTGCAACCTATGAGGAGACCGCTTATGCA 56683  
QY 246 CTCTTATGAGACAGCTCCCACTGCTTATATCACTCACTGCCCCCGAGGATACAGCC 305  
DB 56684 CTCTTATGAGACAGCTCCCACTGCTTATATCACTCACTGCCCCCGAGGATACAGCC 56743  
QY 306 AACCTGTCCAGGGGTATGCACTGCTTATGATATCAACCACTGCTATGACCAACCA 365  
DB 56744 AACCTGTCCAGGGGTATGCACTGCTTATGATATCAACCACTGCTATGACCAACCA 56803  
QY 366 CCCAGGCTCTTATGCACTGCTTATGATATCAACCACTGCTATGACCAACCA 425  
DB 56804 CCCAGGCTCTTATGCACTGCTTATGATATCAACCACTGCTATGACCAACCA 56863  
QY 426 GCGAGAGCGAGGACGACGCTGCACTCAAGACCGAGGTGGAACCAAGCCCACTGAGA 485  
DB 56864 GCGAGAGCGAGGACGACGCTGCACTCAAGACCGAGGTGGAACCAAGCCCACTGAGA 56923  
QY 486 CTATGTCACCTCAATCTAGACAGGAGGTTTACACAGCCAGCCTATGATATGACAGA 545  
DB 56924 CTATGTCACCTCAATCTAGACAGGAGGTTTACACAGCCAGCCTATGATATGACAGA 56983  
QY 546 GTTACTACAGTTATCCCGAGGTAAGTGGAGTCACTCCCATGACCGCATGCTGCACTTC 605  
DB 56984 GTTACTACAGTTATCCCGAGGTAAGTGGAGTCACTCCCATGACCGCATGCTGCACTTC 57043  
QY 606 CATCCCACTCTCTACAGTATTTCTCTACACAGGCGACTATGATATGATCAGAGAGTT 665  
DB 57044 CATCCCACTCTCTACAGTATTTCTCTACACAGGCGACTATGATATGATCAGAGAGTT 57103  
QY 666 ACTCTGACAGAAACCTATGAGGCAACGAGCAGCTATGACAGCAGAGTATGATGTC 725  
DB 57104 ACTCTGACAGAAACCTATGAGGCAACGAGCAGCTATGACAGCAGAGTATGATGTC 57163  
QY 726 AACAAAGCAGCTATGAGGCGAGGCTCTCCATGTTATCCCACTGCACTGATCTTACA 785  
DB 57164 AACAAAGCAGCTATGAGGCGAGGCTCTCCATGTTATCCCACTGCACTGATCTTACA 57223  
QY 786 GCCAAGCTCCAAAGTCAATATGAGCAACAGGAGCAGCTACGAGGCGAGAGTCAATCC 845  
DB 57224 GCCAAGCTCCAAAGTCAATATGAGCAACAGGAGCAGCTACGAGGCGAGAGTCAATCC 57283  
QY 846 GACAGAGCAACCCAGTAGCAGTGGTGTATGAGGAGAGTCTGAGAGATTTTCCGAGC 905  
DB 57284 GACAGAGCAACCCAGTAGCAGTGGTGTATGAGGAGAGTCTGAGAGATTTTCCGAGC 57343  
QY 906 CAGAGAGAAACCGAGCAGTATGAGGCTCTGATTAACGAGGCGAGGAGAGAGGAGTTTG 965  
DB 57344 TAGAGAGAAACCGAGCAGTATGAGGCTCTGATTAACGAGGCGAGGAGAGGAGTTTG 57403  
QY 966 ATGCTGAGAGCATGAGCAGAGTGGGCGAGGAGAGAGAGCGGCTGAGATGGGAGCGTG 1025  
DB 57404 ATGCTGAGAGCATGAGCAGAGTGGGCGAGGAGAGAGAGCGGCTGAGATGGGAGCGTG 57463  
QY 1026 GAGAGCAGAGTGGCTTCAATAGCCTGTGAGCCCAATGATGAGAGCAACAGATCTTGATC 1085  
DB 57464 GAGAGCAGAGTGGCTTCAATAGCCTGTGAGCCCAATGATGAGAGCAACAGATCTTGATC 57523  
QY 1086 TAGGCCCACTGTATGATCCAGATGAGAGCTTGTCAACAGTGAATTTATGTACAAAGAT 1145  
DB 57524 TAGGCCCACTGTATGATCCAGATGAGAGCTTGTCAACAGTGAATTTATGTACAAAGAT 57583

QY 1146 TAAATGACAGTGTACTAGATGATCTGAGAGCTTCTTTAAGAGTGTGGGTGTTA 1205  
DB 57584 TAAATGACAGTGTACTAGATGATCTGAGAGCTTCTTTAAGAGTGTGGGTGTTA 57643  
QY 1206 AGATGACAGAGAACTGGGCAACCCATGATTCACATCTTCTGAGCAAGAGAAACAGAA 1265  
DB 57644 AGATGACAGAGAACTGGGCAACCCATGATTCACATCTTCTGAGCAAGAGAAACAGAA 57703  
QY 1266 AGCCCAAGGAGATGCCAAGTGTCTATGAGAGCCACCCATGAGGCGAGGCTGTG 1325  
DB 57704 AGCCCAAGGAGATGCCAAGTGTCTATGAGAGCTCACTATGCGCAAGCTGCGGTG 57763  
QY 1326 AATGTTTATGAGAGAAATTTTCAAGGAGAGCACTTAAGTCTCTCTGTCTGAGAGA 1385  
DB 57764 AATGTTTATGAGAGAAATTTTCAAGGAGAGCACTTAAGTCTCTCTGTCTGAGAGA 57823  
QY 1386 AGCTTCCATGAGACGTATGCGGGGTGTCTGCCACCCGTGAGGCGAGAGGCTATGCC 1445  
DB 57824 AGCTTCCATGAGACGTATGCGGGGTGTATGCCACCCATGAGGCGAGAGGATGCCAC 57883  
QY 1446 CACCACTCCGTGAGAGGTCCAGAGGCGCAGAGAGTCTGAGGAGACCCATGAGGTGCTG 1505  
DB 57884 CACCACTCCGTGAGAGGTCCAGAGGCGCAGAGAGTCTGAGGAGACCCATGAGGTGCTG 57943  
QY 1506 GAGGCGGTGAGAGATGAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTCCGAGGGA 1565  
DB 57944 GAGGCGGTGAGAGATGAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTCCGAGGGA 58003  
QY 1566 ACCCTCTGAGAGAGAACTGTCAGACCCAGCTGAGAGACTGAGAGTGTCCCATTCGG 1625  
DB 58004 ACCCTCTGAGAGAGAACTGTCAGACCCAGCTGAGAGACTGAGAGTGTCCCATTCGG 58063  
QY 1626 GTTGTGAAACAGAACTTCCCTGAGAGACAGAGTGCACACAGTGTAAAGCCCAAGC 1685  
DB 58064 GTTGTGAAACAGAACTTCCCTGAGAGACAGAGTGCACACAGTGTAAAGCTTCAAGC 58123  
QY 1686 CTGAAGGCTTCTCCCGCACCCTTTCGCGCCCGGGTGTATCTGTGACAGAGTGCC 1745  
DB 58124 CTGAAGGCTTCTCCCGCACCCTTTCGCGCCCGGGTGTATCTGTGACAGAGTGCC 58183  
QY 1746 CTGTGTGACATGCGGGAGAGAGAGTGTCTCATGATGTGTGTGTCTCCGCTGGAATGT 1805  
DB 58184 CTGTGTGACATGCGGGAGAGAGAGTGTCTCATGATGTGTGTGTCTCCGCTGGAATGT 58243  
QY 1806 TCAGAGGTGGCGGTGTGAGACAGAGGTGCTCCGTGTGAGCGCGGAGTGAACCGAG 1865  
DB 58244 TCAGAGGTGGCGGTGTGAGACAGAGGTGCTCCGTGTGAGCGCGGAGTGAACCGAG 58302  
QY 1866 GTGGCTTGTGTGAGAGACAGAGGTGCTCCGTGTGAGCGCGGAGTGAACCGAG 1925  
DB 58303 GTGGCTTGTGTGAGAGACAGAGGTGCTCCGTGTGAGCGCGGAGTGAACCGAG 58361  
QY 1926 AGATGGAGAGAAAGAGAGAGAGCTGTGAGAGACCTGAGAAATATGATTAAGGAGACAC 1985  
DB 58362 AGATGGAGAGAAAGAGAGAGAGCTGTGAGAGACCTGAGAAATATGATTAAGGAGACAC 58421  
QY 1986 GTGAGAGGCGAGAGATGCGGCTCTAGATGAGAGACCCGAGAGCTGATGACTA 2045  
DB 58422 GTGAGAGGCGAGAGATGCGGCTCTAGATGAGAGACCCGAGAGCTGATGACTA 58481  
QY 2046 CCAGATTTATTTTAAACAGAAATGTTTAAATTTAATTTCAATTTAATTTATGTT 2105  
DB 58482 CCAGATTTATTTTAAACAGAAATGTTTAAATTTAATTTCAATTTAATTTATGTT 58541  
QY 2106 GGCACACATATATGATTTATCTGTCTGTATCTTTAGTTTATTTTCACTTTTGTGAAGA 2165  
DB 58542 GGCACACATATATGATTTATCTGTCTGTATCTTTAGTTTATTTTCACTTTTGTGAAGA 58601  
QY 2166 AACATTTAAACAAATTAATGATA 2189  
DB 58602 AACATTTAAACAAATTAATGATA 58625

Tue Feb 22 11:31:10 2005

us-10-791-017a-1.rge

Page 25

Search completed: February 20, 2005, 11:50:21  
Job time : 10045.3 secs

3.4

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 20, 2005, 00:31:32 ; Search time 1159.11 Seconds  
(without alignments)  
12206.093 Million cell updates/sec

Title: US-10-791-017A-1

Perfect score: 2390  
Sequence: 1 agagggagagcgagcgttgcg.....taaaaaaaaaaaaaaaaaa 2390

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1808:\*  
2: geneseqn1908:\*  
3: geneseqn2008:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2390	100.0	2390	13	ADSI6277 Human CDN
2	2388.4	99.9	2390	6	ABK4628 Human CDN
3	2388.4	99.9	2390	6	ABK4628 Human CDN
4	2388.4	99.9	2390	6	ABK4628 Human CDN
5	2388.4	99.9	2390	13	ACN37537 Human Ews
6	2369.4	99.1	2371	2	AAO50643 Human Ews
7	2357.8	98.7	2372	13	ADG86032 Human tum
8	2281.2	95.4	2388	5	AA570647 DNA encod
9	2179.6	91.2	2273	6	AA562623 CDNA sequ
10	2141.8	89.6	2176	6	AA562622 CDNA sequ
11	2033.4	85.1	2177	13	ACN37537 Tumour-as
12	1927.4	80.6	2188	6	AB193833 Mouse lsc
13	1808.4	75.7	1988	10	ADG53506 Human cod
14	1782.8	74.6	2026	13	ADG53506 Human cod
15	1564.4	65.5	1807	13	ADG56333 Human PRO
16	1305	54.6	3309	2	AAQ50646 Human Ews
17	819	34.3	2403	2	AA197870 CDNA enco
18	797.6	33.4	2412	2	AA197869 CDNA enco
19	758.4	31.7	856	13	AB573253 DNA encod
20					ACN37540 Tumour-as

21	707.4	29.6	845	13	ACN37538	ACN37538 Tumour-as
22	653.6	27.3	1462	4	AA126725	AA126725 Human bre
23	570	23.8	2299	11	ACN86697	ACN86697 Breast ca
24	569.4	23.8	665	13	ADG54544	ADG54544 Novel can
25	514.6	21.5	540	4	AA117281	AA117281 Probe #72
26	514.6	21.5	540	4	ABA62106	ABA62106 Human foe
27	514.6	21.5	540	4	AA142050	AA142050 Probe #10
28	514.6	21.5	540	4	ABA29535	ABA29535 Probe #80
29	514.6	21.5	540	4	AAK36327	AAK36327 Human bon
30	514.6	21.5	540	4	AAK10423	AAK10423 Human bra
31	514.6	21.5	540	4	AB536007	AB536007 Human liv
32	514.6	21.5	540	6	AB510394	AB510394 Human gen
33	508.4	21.3	550	9	ACH40478	ACH40478 Human foe
34	488.4	20.4	684	13	ACN37539	ACN37539 Tumour-as
35	480.2	20.1	568	9	ACH22345	ACH22345 Human adu
36	435.4	18.2	468	9	ACH22497	ACH22497 Human end
37	431	18.0	454	3	AAO1659	AAO1659 Human sec
38	419.2	17.5	471	2	AAV86051	AAV86051 EST clone
39	411.6	17.2	505	4	AA117153	AA117153 Probe #70
40	411.6	17.2	505	4	ABA61641	ABA61641 Human foe
41	411.6	17.2	505	4	AA141553	AA141553 Probe #10
42	411.6	17.2	505	4	ABA29300	ABA29300 Probe #77
43	411.6	17.2	505	4	AAK35836	AAK35836 Human bon
44	411.6	17.2	505	4	AAK09942	AAK09942 Human bra
45	411.6	17.2	505	4	AB535546	AB535546 Human liv

## ALIGNMENTS

RESULT 1	ADSI6277	standard; CDNA; 2390 BP.
ID	ADSI6277	
XX	ADSI6277	
XX	ADSI6277	
DT	18-NOV-2004	(first entry)
XX		
DE	Human CDNA encoding the Ewing sarcoma protein Segid 1.	
XX		
KW	human; ss; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne; hypogonadism; androgen-resistance syndrome; testicular feminisation.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	44..2014
FT		/*tag= a
FT		/product= "Ewing sarcoma protein"
XX		
XX	EP1455190-A1.	
XX		
PD	08-SEP-2004.	
XX		
PF	16-FEB-2004; 2004EP-00003422.	
XX		
PR	04-MAR-2003; 2003DE-01009280.	
PR	25-APR-2003; 2003US-0465692P.	
XX		
PA	(SCHD ) SCHERING AG.	
XX		
PI	Obendorf M, Wolf S;	
XX		
DR	WPI; 2004-627861/61.	
DR	F-PSDB; ADSI6278.	
XX		
PT	Determining the hormonal effects of substances, used to identify	
PT	pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,	
PT	from modulating interaction between nuclear receptors and Ewing sarcoma	
XX	protein.	
PS	Claim 7; SEQ ID NO 1; 30pp; German.	
XX		

CC This invention relates to a novel modulators that alter the interaction  
CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
CC as the screening method thereof. Specifically, it refers to determining  
CC and identifying a hormonal effect brought about by test compounds that  
CC modulate either the binding of EWS to the nuclear receptor or the ligand-  
CC induced activity of this receptor. The present invention describes the  
CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
CC Vitamin D, and retinoic acid receptors, most preferably they are androgen  
CC receptors. Accordingly, these modulators may be used in the development  
CC of pharmaceutical compositions that can diagnose and be used to treat  
CC diseases associated with receptor dysfunction such as prostatic cancer,  
CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
CC testicular feminisation. This method provides reliable, sensitive,  
CC simple, inexpensive and rapid assessment of the hormonal effects of these  
CC test compounds. This polynucleotide sequence is the cDNA encoding the  
CC human Ewing sarcoma protein of the invention.

**SQ** Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match	100.0%;	Score 2390;	DB 13;	Length 2390;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2390; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	AGAGGGAGACGGACCGTTGAGAGAACGAGAGAGGAGAGAAATGCGCTCCACGGATTA	60
Db	1	MAAGGAGACGGACCGTTGAGAGAACGAGAGAGGAGAGAAATGCGCTCCACGGATTA	60
QY	61	CAGTACCTATAGCCAAAGCTGCAGCGCAGCAGAGGGCTTACAGTGGCTTACACGGCCAGCCAC	120
Db	61	CAGTACCTATAGCCAAAGCTGCAGCGCAGCAGAGGGCTTACAGTGGCTTACACGGCCAGCCAC	120
QY	121	TCAAAGATATGCACAGACCAACCGCAATATGGGGCAACAAAGCTATGAAACCTATATGACA	180
Db	121	TCAAAGATATGCACAGACCAACCGCAATATGGGGCAACAAAGCTATGAAACCTATATGACA	180
QY	181	GCCCACTGATGTCAAGTATATACCAAGGCTCAGACCACTGCAACCTATGGGAGACCCCTTA	240
Db	181	GCCCACTGATGTCAAGTATATACCAAGGCTCAGACCACTGCAACCTATGGGAGACCCCTTA	240
QY	241	TGCAACTTCTTATGACAGAGCTCCCACTGGTTATACTTCCACTGCCCCCAGGCATA	300
Db	241	TGCAACTTCTTATGACAGAGCTCCCACTGGTTATACTTCCACTGCCCCCAGGCATA	300
QY	301	CAGCCAGCGCTGTCAGAGGGGATATGGCACTGGTGTATATGATACCAACCACTGCTACAGTCA	360
Db	301	CAGCCAGCGCTGTCAGAGGGGATATGGCACTGGTGTATATGATACCAACCACTGCTACAGTCA	360
QY	361	CACCAACCCAGGCGCTCTATGACAGCTCAGTCTGCATATGGCACTCAGCTGCTTATCCAGC	420
Db	361	CACCAACCCAGGCGCTCTATGACAGCTCAGTCTGCATATGGCACTCAGCTGCTTATCCAGC	420
QY	421	CTATGGGCGAGCCGACGACCACTGCACTTAACAAACCGAGAGATGAAACAAAGCCAC	480
Db	421	CTATGGGCGAGCCGACGACCACTGCACTTAACAAACCGAGAGATGAAACAAAGCCAC	480
QY	481	TGAGACTATCAACCTCAATCTTAGCAACAGGGGTTTCAACACAGCCAGCTTAGATATGG	540
Db	481	TGAGACTATCAACCTCAATCTTAGCAACAGGGGTTTCAACACAGCCAGCTTAGATATGG	540
QY	541	ACAGAGTAACTACAGTTATCCCAAGGTACTTGGAGCTTACCCATGCAAGCACTCATCTGC	600
Db	541	ACAGAGTAACTACAGTTATCCCAAGGTACTTGGAGCTTACCCATGCAAGCACTCATCTGC	600
QY	601	ACCTCATCTTACCCCTCCATCAACAGGTAATCTCTTACAGCGGCACTATGATATGATCAGAG	660
Db	601	ACCTCATCTTACCCCTCCATCAACAGGTAATCTCTTACAGCGGCACTATGATATGATCAGAG	660
QY	661	CAGTTACTCTCAGCAGAAACCTATGGGCAACCGACAGCTATATGACAGCAGATAGCTTA	720
Db	661	CAGTTACTCTCAGCAGAAACCTATGGGCAACCGACAGCTATATGACAGCAGATAGCTTA	720
QY	721	TGGTCAACAAGCAGTTATGGGAGCAGAGCTCCCACTATGTTACCAACCCAACTGGATC	780

Db	721	TGCTCAACAAGCAGTAATG3GCGAGAGCTCCCACTAGATTACCAACCCCAACTGATC	780
QY	781	CTAAGCCCAAGCTTCCAAGTCAATATATAGCCCAACAGACGACTACCGGCGACGAAATTC	840
Db	781	CTACAGCCCAAGCTTCCAAGTCAATATATAGCCCAACAGACGACTACCGGCGACGAAATTC	840
QY	841	ATTCCGACAGGACCAACCCCACTAGTACATG3GCTGTTTATATG3GACAGGATCTCTGAGGATTTTC	900
Db	841	ATTCCGACAGGACCAACCCCACTAGTACATG3GCTGTTTATATG3GACAGGATCTCTGAGGATTTTC	900
QY	901	CGACCCAGAGAGAACCCGAGCATGATG3GCCCTTGATTAACCG3G3GACG3G3AAGAG3G3G	960
Db	901	CGACCCAGAGAGAACCCGAGCATGATG3GCCCTTGATTAACCG3G3GACG3G3AAGAG3G3G	960
QY	961	ATTATGATCTGTGAGAGCATGAGCAGAGGTG3GCG3G3GAGAGAGACGCGGTG3AATG3G3CAG	1020
Db	961	ATTATGATCTGTGAGAGCATGAGCAGAGGTG3GCG3G3GAGAGAGACGCGGTG3AATG3G3CAG	1020
QY	1021	CGCTG3GAGGAGGAGGTG3GCTTCAATTAAGCCTG3G3GACCCATGGATGAAGGACCAAGATCT	1080
Db	1021	CGCTG3GAGGAGGAGGTG3GCTTCAATTAAGCCTG3G3GACCCATGGATGAAGGACCAAGATCT	1080
QY	1081	TGATCTAG3GCCCACTCTGTAGTCCAGATGGAAGACTGACAAACAGTGCMAATTAATGTACA	1140
Db	1081	TGATCTAG3GCCCACTCTGTAGTCCAGATGGAAGACTGACAAACAGTGCMAATTAATGTACA	1140
QY	1141	AGGATTAATATGACAGTGTGACTCTTAATGATCTG3GACACTTCTTTAAGCATGTG3GGGT	1200
Db	1141	AGGATTAATATGACAGTGTGACTCTTAATGATCTG3GACACTTCTTTAAGCATGTGTG3GGGT	1200
QY	1201	TGTTAATATGAAACAAGAGAACTG3G3GCAACCCATGATCCATCATCTGATCTG3GACAAAGAAAC	1260
Db	1201	TGTTAATATGAAACAAGAGAACTG3G3GCAACCCATGATCCATCATCTGATCTG3GACAAAGAAAC	1260
QY	1261	AGGAAAGCCCAAGAGCCGATGCAACAGTGTCTTAATGAAGACCAACCACTG3GCAAG3GTGC	1320
Db	1261	AGGAAAGCCCAAGAGCCGATGCAACAGTGTCTTAATGAAGACCAACCACTG3GCAAG3GTGC	1320
QY	1321	CGTGAATATGTTTATGATG3GAAAGATTTTCAAG3GAGCAAACTTAAGTCTTCCCTGTGCTCG	1380
Db	1321	CGTGAATATGTTTATGATG3GAAAGATTTTCAAG3GAGCAAACTTAAGTCTTCCCTGTGCTCG	1380
QY	1381	GAAAGACCTTCCCAATGAACAGTATG3G3G3GTGTCTGCCAACCCTGTAG3G3CAGAGGCAT	1440
Db	1381	GAAAGACCTTCCCAATGAACAGTATG3G3G3GTGTCTGCCAACCCTGTAG3G3CAGAGGCAT	1440
QY	1441	GCCACCAACCACTCCGTG3AGGTCCAGAGAGGCCCAAGAGTCTTGTG3G3GACCCATG3G3TGC	1500
Db	1441	GCCACCAACCACTCCGTG3AGGTCCAGAGAGGCCCAAGAGTCTTGTG3G3GACCCATG3G3TGC	1500
QY	1501	CATG3GAGGCCGTGTG3AGAGATGAGAGAGGCTTCTCTCAAGAGAGACCCCG3G3TTCGCG	1560
Db	1501	CATG3GAGGCCGTGTG3AGAGATGAGAGAGGCTTCTCTCAAGAGAGACCCCG3G3TTCGCG	1560
QY	1561	AGGGAACCCCTCTGTG3AGAGAGGAACCTTCCAGACCCGAGCTGTG3AATCTGTG3CATTC	1620
Db	1561	AGGGAACCCCTCTGTG3AGAGAGGAACCTTCCAGACCCGAGCTGTG3AATCTGTG3CATTC	1620
QY	1621	TCCG3GTTTGTG3AACAAGAACTTCCGTG3GAGAAACAGAGTGCACAACAGTGTAAAGCCCC	1680
Db	1621	TCCG3GTTTGTG3AACAAGAACTTCCGTG3GAGAAACAGAGTGCACAACAGTGTAAAGCCCC	1680
QY	1681	AAAGCCTTAAGAGCTTCTCTCCGCCACCTTTCGCGCCCG3G3GTGTATCTGTG3CAGAGG	1740
Db	1681	AAAGCCTTAAGAGCTTCTCTCCGCCACCTTTCGCGCCCG3G3GTGTATCTGTG3CAGAGG	1740
QY	1741	TG3GCCCTGTG3TGTG3CATG3G3G3AGAGAAAGTGTG3CTCATGTATGTGTG3TCCCGGTG3	1800
Db	1741	TG3GCCCTGTG3TGTG3CATG3G3G3AGAGAAAGTGTG3CTCATGTATGTGTG3TCCCGGTG3	1800
QY	1801	AATGTTCACAGAGT3GCGGTGTG3GACAGAGTGTG3GCTCCGTGTG3G3G3G3CATATGA	1860
Db	1801	AATGTTCACAGAGT3GCGGTGTG3GACAGAGTGTG3GCTCCGTGTG3G3G3G3CATATGA	1860



Db	361	CACCAACCAGGCGCTCTCTATGCACTCACTGTGCATATGGCACTACGCTCTGTTATCCAGC	420
Qy	421	CTATGGGAGAGCCAGCAGCCACTGCACTTACAAAGCCGAGGATGAAACAAAGCCAC	480
Db	421	CTATGGGAGAGCCAGCAGCCACTGCACTTACAAAGCCGAGGATGAAACAAAGCCAC	480
Qy	481	TGAACCTGTCAACTCTCAATTTAGCAGAGGGGGTTTCAACAGGCCAGCCCTAGATATGG	540
Db	481	TGAACCTGTCAACTCTCAATTTAGCAGAGGGGGTTTCAACAGGCCAGCCCTAGATATGG	540
Qy	541	ACAGAGTAATACTACAGTTATCCCAAGGTACTCTGGAGAGCTAACCCCATGCACTGC	600
Db	541	ACAGAGTAATACTACAGTTATCCCAAGGTACTCTGGAGAGCTAACCCCATGCACTGC	600
Qy	601	ACCTTCATCTTACCCCTCCTACAGCTATTCCTTACACAGCCGACTAGTTATGATCAAG	660
Db	601	ACCTTCATCTTACCCCTCCTACAGCTATTCCTTACACAGCCGACTAGTTATGATCAAG	660
Qy	661	CAGTTACTCTCAGCAGAACACTTATGGGCAACCGAGAGCTATGACAGCAGATTACTA	720
Db	661	CAGTTACTCTCAGCAGAACACTTATGGGCAACCGAGAGCTATGACAGCAGATTACTA	720
Qy	721	TGGTCAATAAAGCAGCTATGGGAGCAGCCTCCCACTAGTTACCAACCCCAACTGATC	780
Db	721	TGGTCAATAAAGCAGCTATGGGAGCAGCCTCCCACTAGTTACCAACCCCAACTGATC	780
Qy	781	CTACAGCCAGCTCCCAAGTCAATATAGCCAAAGAGCAGCAGCTACGGGCGAGAGTTC	840
Db	781	CTACAGCCAGCTCCCAAGTCAATATAGCCAAAGAGCAGCAGCTACGGGCGAGAGTTC	840
Qy	841	ATTCGCAAGAGCAACCCCAAGTACATGGGTGTTTATGGGCAAGAGTCTGGAGATTTTC	900
Db	841	ATTCGCAAGAGCAACCCCAAGTACATGGGTGTTTATGGGCAAGAGTCTGGAGATTTTC	900
Qy	901	CGGACAGAGAGAAACCGGAGCATGAGTGGCCCTGATACCGGGGCGAGGGGAAAGGGGG	960
Db	901	CGGACAGAGAGAAACCGGAGCATGAGTGGCCCTGATACCGGGGCGAGGGGAAAGGGGG	960
Qy	961	ATTGATGCTGAGAGCACTAGCAGAGTGGGGCGGGAGAGAGACGCGGTGAAATGGGAG	1020
Db	961	ATTGATGCTGAGAGCACTAGCAGAGTGGGGCGGGAGAGAGACGCGGTGAAATGGGAG	1020
Qy	1021	CGCTGGAAGCGAGGTGCTTCAATTAAGCTGTGTGAACCCATGATGAAGACCAATCT	1080
Db	1021	CGCTGGAAGCGAGGTGCTTCAATTAAGCTGTGTGAACCCATGATGAAGACCAATCT	1080
Qy	1081	TGATCTAGGCGCACTGTATGATCCAGATGAAGACTGTACAAAGTGCATTTATATGACA	1140
Db	1081	TGATCTAGGCGCACTGTATGATCCAGATGAAGACTGTACAAAGTGCATTTATATGACA	1140
Qy	1141	AGGATTTAAATGACAGTGTGACTCTATGATGATCTGGCAGACTTCTTTAAGCAGTGGGGT	1200
Db	1141	AGGATTTAAATGACAGTGTGACTCTATGATGATCTGGCAGACTTCTTTAAGCAGTGGGGT	1200
Qy	1201	TGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCAATCTTACCTGGAACAAAGAAAC	1260
Db	1201	TGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCAATCTTACCTGGAACAAAGAAAC	1260
Qy	1261	AGGAAAGCCCAAGGCGATGCAAGTGTCCATATGAACCAACCACTGCAAGGCTGC	1320
Db	1261	AGGAAAGCCCAAGGCGATGCAAGTGTCCATATGAACCAACCACTGCAAGGCTGC	1320
Qy	1321	CGTGAATGCTTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTGC	1380
Db	1321	CGTGAATGCTTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTGC	1380
Qy	1381	GAAAGAGCCTCCAAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAAGGCAAGGCAT	1440
Db	1381	GAAAGAGCCTCCAAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAAGGCAAGGCAT	1440
Qy	1441	GCCACCAACCACTTCCTGAGAGTTCAGAGAGGCCAAGAGGTCTCTGGGGGACCAATGGGTGC	1500
Db	1441	GCCACCAACCACTTCCTGAGAGTTCAGAGAGGCCAAGAGGTCTCTGGGGGACCAATGGGTGC	1500

Dp	1441	GCCACCAACCACTCCGTGGAGGTCCAGGAGGCCCCAGGAGGTCTCGGGGGAACTCATGTGGCTG	1500
Qy	1501	CATGGGAGGCGCCGTGGAGAGATAGAGAGGCTTCCCTCCAAGAGAACCCCGGGGTTCCG	1560
Dp	1501	CATGGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCTCCAAGAGAACCCCGGGGTTCCG	1560
Qy	1561	AGGGAACCCCTCTGGAGAGAAAGTCCAGCAACCGAGCTGGAGACTGGACAGTGTCCCA	1620
Dp	1561	AGGGAACCCCTCTGGAGAGAAAGTCCAGCAACCGAGCTGGAGACTGGAGACTGGATCCCA	1620
Qy	1621	TCCGGGTTGTGAAACCAAGAACTTCCGCTGAGAACAGAGTGCACCAAGTAAAGGCC	1680
Dp	1621	TCCGGGTTGTGAAACCAAGAACTTCCGCTGAGAACAGAGTGCACCAAGTAAAGGCC	1680
Qy	1681	AAACCTGGAAGGCTTCTCCCGCACCCCTTCCGCCCCCGGGTGTGATGTGTGCAAGG	1740
Dp	1681	AAACCTGGAAGGCTTCTCCCGCACCCCTTCCGCCCCCGGGTGTGATGTGTGCAAGG	1740
Qy	1741	TGGCCCTGTGTGAGCAATGCGGGAGAGAGAGGTGCGCTCATGATGTGTGTGCTCCGCTG	1800
Dp	1741	TGGCCCTGTGTGAGCAATGCGGGAGAGAGAGGTGCGCTCATGATGTGTGTGCTCCGCTG	1800
Qy	1801	AATGTTCAAGAGTGGCCGTGGTGGAGACAAGAGTGTGCTTCCGTGTGGCCGGGCAATGA	1860
Dp	1801	AATGTTCAAGAGTGGCCGTGGTGGAGACAAGAGTGTGCTTCCGTGTGGCCGGGCAATGA	1860
Qy	1861	CCGAGGTGTGCTTGTGTGAGAGAAACAAGTGTGCGCTCTGGGGGGCCCCCTGTGACTTGTAT	1920
Dp	1861	CCGAGGTGTGCTTGTGTGAGAGAAACAAGTGTGCGCTCTGGGGGGCCCCCTGTGACTTGTAT	1920
Qy	1921	GGAACAGATGGGAGAAAGAGAGAGGACGCTGAGAGCACTGGAATAATGTGATAAGCGA	1980
Dp	1921	GGAACAGATGGGAGAAAGAGAGAGGACGCTGAGAGCACTGGAATAATGTGATAAGCGA	1980
Qy	1981	GCACCGTCAGAGGCGCAGAGATCGGCCCTTACTAGATCAGAGACCCCGCAGAGCTGCATT	2040
Dp	1981	GCACCGTCAGAGGCGCAGAGATCGGCCCTTACTAGATCAGAGACCCCGCAGAGCTGCATT	2040
Qy	2041	GACTACCAAGATTATTTTTTAAACAGAAATGTTTAAATTATTAATTCGATATTATA	2100
Dp	2041	GACTACCAAGATTATTTTTTAAACAGAAATGTTTAAATTATTAATTCGATATTATA	2100
Qy	2101	ATGTTGGCAACAATATATGATTTATCTGTGCTGACTTATAGTATTTTTCACCATTTGT	2160
Dp	2101	ATGTTGGCAACAATATATGATTTATCTGTGCTGACTTATAGTATTTTTCACCATTTGT	2160
Qy	2161	GAAAGAACATTAATAACAAGTTAAATGTGATGTGCGAGTTTTTTTTCTTCCTCTTT	2220
Dp	2161	GAAAGAACATTAATAACAAGTTAAATGTGATGTGCGAGTTTTTTTTCTTCCTCTTT	2220
Qy	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGACATGCTCAGATCAAT	2280
Dp	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGACATGCTCAGATCAAT	2280
Qy	2281	GTGAGGAACCAAGAGGCGCTTAACTGTAACAATGTCAATGTCAATGTGTGATTTTTTTT	2340
Dp	2281	GTGAGGAACCAAGAGGCGCTTAACTGTAACAATGTCAATGTCAATGTGTGATTTTTTTT	2340
Qy	2341	TTTTTTTAAATTAATTCCAATGTTTAAATAAAAAAAAAAAAAAAAAAAAAA	2390
Dp	2341	TTTTTTTAAATTAATTCCAATGTTTAAATAAAAAAAAAAAAAAAAAAAAAA	2390
RESULT 3			
ABK64822			
ID	ABK64822 standard; DNA; 2390 BP.		
AC	ABK64822;		
XX			
XX	18-JUN-2002 (first entry)		
DT			
XX			
DE	Human benign prostatic hyperplasia gene #717.		
XX			



KM Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
 XX Homo sapiens.  
 OS  
 XX WO200212440-A2.  
 XX  
 XX 14-FEB-2002.  
 XX  
 XX 07-AUG-2001; 2001MO-US024708.  
 XX  
 XX 07-AUG-2000; 2000US-0223323P.  
 XX 05-JUN-2001; 2001US-00873319.  
 XX  
 XX (GENE-) GENE LOGIC INC.  
 XX (NISB) JAPAN TOBACCO INC.  
 XX  
 XX Munger WE, Kulikarni P, Getzenberg RH, Waga I, Yamamoto J;  
 DR WPI; 2002-257476/30.  
 XX  
 XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 XX  
 XX Disclosure; Page 399-400; 444pp; English.  
 PS  
 XX The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 CC  
 XX  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 Query Match 99.9%; Score 2388.4; DB 6; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 241 TGCACTTCTTATGACAGGCTTCCAGTGTATTAATCTCAACTGCCCCCAGGAGATA 300  
 Qy 301 CAGCCAGCCTGTTCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCACTAC 360  
 Db 301 CAGCCAGCCTGTTCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCACTAC 360  
 Qy 361 CAGCCAGCCTGTTCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCACTAC 420  
 Db 361 CAGCCAGCCTGTTCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCACTAC 420  
 Qy 421 CTAATGGGAGAGCCAGGAGCCAGTCACTGATCAATCAAGGAGATGAAACAGGCTAC 480  
 Db 421 CTAATGGGAGAGCCAGGAGCCAGTCACTGATCAATCAAGGAGATGAAACAGGCTAC 480  
 Qy 481 TGAGACTAGTCAACTCTAATCTAGCAAGGGGGTTACACAGCCAGCCTAGATATATGG 540  
 Db 481 TGAGACTAGTCAACTCTAATCTAGCAAGGGGGTTACACAGCCAGCCTAGATATATGG 540  
 Qy 541 ACAGAGTAACTACAGTTATCCCGAGTACCTGGAGCTACCCATGCAAGTCACTGC 600  
 Db 541 ACAGAGTAACTACAGTTATCCCGAGTACCTGGAGCTACCCATGCAAGTCACTGC 600  
 Qy 601 ACCTCATCTCTACCTCCCTACAGAGCTAATCTCTACACAGCCGACTAGTATATGATCAGAG 660  
 Db 601 ACCTCATCTCTACCTCCCTACAGAGCTAATCTCTACACAGCCGACTAGTATATGATCAGAG 660  
 Qy 661 CAGTTACTCTCAGCAGAACCTATATGGGCAACGAGCAGCTATGACAGAGATAGTACTA 720  
 Db 661 CAGTTACTCTCAGCAGAACCTATATGGGCAACGAGCAGCTATGACAGAGATAGTACTA 720  
 Qy 721 TGGTCAACAAAGAGTATATGGGAGAGCTTCCCACTAGTTACCAACCCCAACTGATATC 780  
 Db 721 TGGTCAACAAAGAGTATATGGGAGAGCTTCCCACTAGTTACCAACCCCAACTGATATC 780  
 Qy 781 CTACAGCCAGAGCTCAAGTCAATATATGCAACAGAGCAGTACAGGAGAGAGTTC 840  
 Db 781 CTACAGCCAGAGCTCAAGTCAATATATGCAACAGAGCAGTACAGGAGAGAGTTC 840  
 Qy 841 ATTCCGACAGAGCAACCCAGTATGATGAGTGTATATGGGAGAGTCTGAGAGATTTTC 900  
 Db 841 ATTCCGACAGAGCAACCCAGTATGATGAGTGTATATGGGAGAGTCTGAGAGATTTTC 900  
 Qy 901 CGGACCAAGAGAAACCGAGATATGAGTGGCTCTGATTAACCGGGGAGAGGAGAGGAG 960  
 Db 901 CGGACCAAGAGAAACCGAGATATGAGTGGCTCTGATTAACCGGGGAGAGGAGAGGAG 960  
 Qy 961 ATTGATGATGAGAGAGAGAGAGAGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 Db 961 ATTGATGATGAGAGAGAGAGAGAGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 Qy 1021 CGCTGAGAGAGAGAGTGGCTTCAATATAGCCTGATGAGCCATGATGAGAGAGAGATCT 1080  
 Db 1021 CGCTGAGAGAGAGAGTGGCTTCAATATAGCCTGATGAGCCATGATGAGAGAGAGATCT 1080  
 Qy 1081 TGATCTAGGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 1081 TGATCTAGGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Qy 1141 AGGATTAATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 Db 1141 AGGATTAATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 Qy 1201 TGTTAAGATGAAACAAGAGATGAGGCAACCATGATGATGATGATGATGATGATGATGAT 1260  
 Db 1201 TGTTAAGATGAAACAAGAGATGAGGCAACCATGATGATGATGATGATGATGATGATGAT 1260  
 Qy 1261 AGGAAAGCCCAAGGAGATGAGGCAAGTGTCTATGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 1261 AGGAAAGCCCAAGGAGATGAGGCAAGTGTCTATGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Qy 1321 CGTGAATGATGATGAGAGAGAGATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

Db	1321	CGTGGATGGTTTGATGGAAAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCG	1380
Qy	1381	GAAGAAAGCTTCMAATGMAACAGTATGCGGGGTGATCTGGCAACCCCGTGAAGGACAGAGCAT	1440
Db	1381	GAAGAAAGCTTCMAATGMAACAGTATGCGGGGTGATCTGGCAACCCCGTGAAGGACAGAGCAT	1440
Qy	1441	GCCACCAACCATTCCTGTGAAGGTCCAGAGAGGCCAGAGAGTCTCTGGGGGACCAATGGGTGCG	1500
Db	1441	GCCACCAACCATTCCTGTGAAGGTCCAGAGAGGCCAGAGAGTCTCTGGGGGACCAATGGGTGCG	1500
Qy	1501	CATAGGAGAGCGGTGAGAGAGATAGAGAGAGCTTCCTCCAAGAGGACCCCGAGGTTCCCG	1560
Db	1501	CATAGGAGAGCGGTGAGAGAGATAGAGAGAGCTTCCTCCAAGAGGACCCCGAGGTTCCCG	1560
Qy	1561	AGGGAAACCCCTCTGGAGGAGGAAACGTCCAGACCCGAGCTGAGACTGGGCAGTGTCCCAA	1620
Db	1561	AGGGAAACCCCTCTGGAGGAGGAAACGTCCAGACCCGAGCTGAGACTGGGCAGTGTCCCAA	1620
Qy	1621	TCCGGGTTGTGGAAGAACGAACCTTGCGTGGAGAACAGATGCAACCAAGTAAAGGCCCC	1680
Db	1621	TCCGGGTTGTGGAAGAACGAACCTTGCGTGGAGAACAGATGCAACCAAGTAAAGGCCCC	1680
Qy	1681	AAAGCTGAAAGGCTTCTCCGCCACCCCTTTCGGCCCCGGGATGTGATCTGTGCAGAGG	1740
Db	1681	AAAGCTGAAAGGCTTCTCCGCCACCCCTTTCGGCCCCGGGATGTGATCTGTGCAGAGG	1740
Qy	1741	TGGCCTCTGTGTGCATGCGGGGAGGAAGAGTGTGCTCATGATCTGTGTGTCTCCCGTGG	1800
Db	1741	TGGCCTCTGTGTGCATGCGGGGAGGAAGAGTGTGCTCATGATCTGTGTGTCTCCCGTGG	1800
Qy	1801	AATGTTCAGAGGTGCGCGTGTGTGAGACAGAGTGTGCTTCGCTGGTGTGCGGGGACATGGA	1860
Db	1801	AATGTTCAGAGGTGCGCGTGTGTGAGACAGAGTGTGCTTCGCTGGTGTGCGGGGACATGGA	1860
Qy	1861	CCGAGGTGACTTTGTGTGAGGAAGACGAGGTGCGCTTGAGGGGCCCTCTGACCTTTGAT	1920
Db	1861	CCGAGGTGACTTTGTGTGAGGAAGACGAGGTGCGCTTGAGGGGCCCTCTGACCTTTGAT	1920
Qy	1921	GGAACAGATGGGAGGAAAGAAAGAGAGACGTGAGAGACCTGAGAAAATAGATAAAGGCGA	1980
Db	1921	GGAACAGATGGGAGGAAAGAAAGAGAGACGTGAGAGACCTGAGAAAATAGATAAAGGCGA	1980
Qy	1981	GCACCGTACAGAGCGCAGAGAGATCGGCCCCCTACTAGATGACAGAGACCCGCAAGGTGCATT	2040
Db	1981	GCACCGTACAGAGCGCAGAGAGATCGGCCCCCTACTAGATGACAGAGACCCGCAAGGTGCATT	2040
Qy	2041	GACTACCAAGATTTATTTTTTAAACAGAAAAATGTTTTAAATTTAATTTCCATATTATA	2100
Db	2041	GACTACCAAGATTTATTTTTTAAACAGAAAAATGTTTTAAATTTAATTTCCATATTATA	2100
Qy	2101	ATGTGTGACCAACAACTATAGATTTCTCTGTCTGTACTTTAGATTTTTCACCAATTGT	2160
Db	2101	ATGTGTGACCAACAACTATAGATTTCTCTGTCTGTACTTTAGATTTTTCACCAATTGT	2160
Qy	2161	GAAGAAACATTAAACCAAGTTAAATGTAGTGTGCGAGTTTTTTTTCTTCTCTCTT	2220
Db	2161	GAAGAAACATTAAACCAAGTTAAATGTAGTGTGCGAGTTTTTTTTCTTCTCTCTT	2220
Qy	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTTGGAGAGATCTCAGATACAT	2280
Db	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTTGGAGATCTCAGATACAT	2280
Qy	2281	GTGAGAACCAAGAGGCGCTCTTAACTGTACAAATGTTCATGTGTGATGTTTTTTTT	2340
Db	2281	GTGAGAACCAAGAGGCGCTCTTAACTGTACAAATGTTCATGTGTGATGTTTTTTTT	2340
Qy	2341	TTTTTTTAAATAAATTCCAAATGTTTATATAAAAATTTTTTTTTTTTT 2390	
Db	2341	TTTTTTTAAATAAATTCCAAATGTTTATATAAAAATTTTTTTTTTTTT 2390	

RESULT 4  
ABN97274

ID		ABN97274 standard; DNA; 2390 BP.
XX	AC	
XX	ABN97274;	
DT	13-AUG-2002	(first entry)
XX		
DE	Gene #3772 used to diagnose liver cancer.	
KW	Gene; liver cancer; dr; hepatocellular carcinoma; hepatotropic;	
KM	metastatic liver tumour; cytostatic; expression profile; disease state;	
KX	disease progression; drug toxicity; drug efficacy; drug metabolism.	
OS	Homo sapiens.	
PN	WO200229103-A2.	
PD	11-APR-2002.	
PF	02-OCT-2001; 2001WO-US030589.	
PR	02-OCT-2000; 2000US-0237054P.	
PA	(GENE-) GENE LOGIC INC.	
XX		
PI	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;	
DR	WPI; 2002-426119/45.	
PT	Diagnosing and detecting the progression of liver cancer, hepatocellular	
PT	carcinoma or metastatic liver tumor in a patient, involves detecting the	
XX	level of expression of two or more genes in a liver tissue sample.	
PS	Claim 1; SEQ ID NO 3772; 298bp; English.	
CC	The invention relates to a novel method for diagnosing and detecting the	
CC	progression of liver cancer, hepatocellular carcinoma or metastatic liver	
CC	tumour in a patient, and differentiating metastatic liver cancer from	
CC	hepatocellular carcinoma in a patient, involving detecting the level of	
CC	expression of two or more genes represented in ABN93503-ABN97455 in a	
CC	tissue sample. The method of the invention has hepatotrophic, and	
CC	cytostatic activity. The method is useful for diagnosing and detecting	
CC	the progression of liver cancer, hepatocellular carcinoma and metastatic	
CC	liver carcinoma in a patient. The method is useful for identifying	
CC	expression profiles which serve as useful diagnostic markers as well as	
CC	markers that can be used to monitor disease states, disease progression,	
CC	drug toxicity, drug efficacy and drug metabolism. Note: The sequence data	
CC	for this patent did not form part of the printed specification, but was	
CC	obtained in electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences	
SQ	Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;	
Query Match	99.9%; Score 2388.4; DB 6; Length 2390;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2389; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
Dn	1 AGAGGGAGACCGAAGCTTGAGAGAACAAGAGAGAAATGGCGTTCACGAGTTA 60	
	1 AGAGGGAGACCGAAGCTTGAGAGAACAAGAGAGAAATGGCGTTCACGAGTTA 60	
Dn	61 CAGTAACCTATAGCCAAAGGTGCAGGCCAGCAGGCTTACAGTCCTTACACCGCCAGCCAC 120	
	61 CAGTAACCTATAGCCAAAGGTGCAGGCCAGCAGGCTTACAGTCCTTACACCGCCAGCCAC 120	
Dn	121 TCAGGATATGCAAGACACCAGGCGCATATGGGCAACAAAGCTATGGAACTTATGGACA 180	
	121 TCAGGATATGCAAGACACCAGGCGCATATGGGCAACAAAGCTATGGAACTTATGGACA 180	
Dn	181 GCCACATGATGTCAGCTATACCAAGGCTCAGAACCACTGCAACTCTATATGGGACGAGCCCTTA 240	
	181 GCCACATGATGTCAGCTATACCAAGGCTCAGAACCACTGCAACTCTATATGGGACGAGCCCTTA 240	
Dn	241 TGCAACTTCTATGAGACAGCTCCCACTGTATTATATACTCCAATGCCCCCAGGCATA 300	
	241 TGCAACTTCTATGAGACAGCTCCCACTGTATTATATACTCCAATGCCCCCAGGCATA 300	

Db 241 TGAACCTTCTTATGAGACGCTCCCACTGGTTATCTCACTCCACCTCCCGGAGATG 300  
QY 301 CAGCCAGCCTGTCAGGGGTATGCACTGGTGTCTTATGATACCACTGCTCACTAGCTAC 360  
Db 301 CAGCCAGCCTGTCAGGGGTATGCACTGGTGTCTTATGATACCACTGCTCACTAGCTAC 360  
QY 361 CAGCCAGCCTGTCAGGGGTATGCACTGGTGTCTTATGATACCACTGCTCACTAGCTAC 420  
Db 361 CAGCCAGCCTGTCAGGGGTATGCACTGGTGTCTTATGATACCACTGCTCACTAGCTAC 420  
QY 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
Db 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
QY 481 TGAAGTACTGCACTCTCAATCTAGCAAGGGGGTTTACACAGCCAGCCTAGATATG 540  
Db 481 TGAAGTACTGCACTCTCAATCTAGCAAGGGGGTTTACACAGCCAGCCTAGATATG 540  
QY 541 ACAGAGTAACTAGAGTTATCCCAAGTAACTTGGAGAGTACCCCACTGAGCAGCTGTC 600  
Db 541 ACAGAGTAACTAGAGTTATCCCAAGTAACTTGGAGAGTACCCCACTGAGCAGCTGTC 600  
QY 601 ACCTCCATCTTACCTCTCTACAGCTATCTCTCTACAGCAGCAGCAGCTATGATGAG 660  
Db 601 ACCTCCATCTTACCTCTCTACAGCTATCTCTCTACAGCAGCAGCAGCTATGATGAG 660  
QY 661 CAGTTACTCTCAGCAGAACCTATGGGCAACGAGCAGCTATGAGCAGCAGTATGCTA 720  
Db 661 CAGTTACTCTCAGCAGAACCTATGGGCAACGAGCAGCTATGAGCAGCAGTATGCTA 720  
QY 721 TGGTCAACAAAGCAGCTATGGGCAAGCCTCCCACTAGTTACCCCACTGATC 780  
Db 721 TGGTCAACAAAGCAGCTATGGGCAAGCCTCCCACTAGTTACCCCACTGATC 780  
QY 781 CTACAGCAGCAGCTCAAGTCAATATAGCAAGAGCAGCAGCTAGCAGCAGCAGCTG 840  
Db 781 CTACAGCAGCAGCTCAAGTCAATATAGCAAGAGCAGCAGCTAGCAGCAGCAGCTG 840  
QY 841 ATTCCGACAGGACACCCAGTAGAGTGGTGTATAGGCAAGAGTCTGAGAGATTTTC 900  
Db 841 ATTCCGACAGGACACCCAGTAGAGTGGTGTATAGGCAAGAGTCTGAGAGATTTTC 900  
QY 901 CGGACCAAGAGAGAACCGGAGCAGTAGTGGCTTGATTAACCGGGGCAAGAGAGAGG 960  
Db 901 CGGACCAAGAGAGAACCGGAGCAGTAGTGGCTTGATTAACCGGGGCAAGAGAGAGG 960  
QY 961 ATTGATCGTGAAGCAGTAGAGTGGGCGGGGAGAGAGCGGGTGGAAATGGGCG 1020  
Db 961 ATTGATCGTGAAGCAGTAGAGTGGGCGGGGAGAGAGCGGGTGGAAATGGGCG 1020  
QY 1021 CGGTGAGAGCGAGTGGCTTCAATAAGCCTGTGTGAGCCCATGATGAGAGCAAGATCT 1080  
Db 1021 CGGTGAGAGCGAGTGGCTTCAATAAGCCTGTGTGAGCCCATGATGAGAGCAAGATCT 1080  
QY 1081 TGAATCTAGGCGCCCTGTGATCTCAATGAGACTGTGACACAGTGCATTTATGTA 1140  
Db 1081 TGAATCTAGGCGCCCTGTGATCTCAATGAGACTGTGACACAGTGCATTTATGTA 1140  
QY 1141 AGGATTAATGACAGTGTGACTCTAGATGATCTGGGAGACTTTCTTTAAGAGTGGGG 1200  
Db 1141 AGGATTAATGACAGTGTGACTCTAGATGATCTGGGAGACTTTCTTTAAGAGTGGGG 1200  
QY 1201 TGTTAAGATGAGCAAGAGAACTGGGCAACCATGATCTCACTGAGCAAGAGAA 1260  
Db 1201 TGTTAAGATGAGCAAGAGAACTGGGCAACCATGATCTCACTGAGCAAGAGAA 1260  
QY 1261 AGGAAAGCCCAAGGGGATGCGCAGTGTCTTATGAGAGCCCACTGCGCAAGGCTGC 1320  
Db 1261 AGGAAAGCCCAAGGGGATGCGCAGTGTCTTATGAGAGCCCACTGCGCAAGGCTGC 1320  
QY 1321 CGTGAATGGTTGATGGAGAAATTTTCAAGGAGCAACTTAAAGTCCCTGCTCG 1380

Db 1321 CGTGAATGGTTGATGGAGAAATTTTCAAGGAGCAACTTAAAGTCCCTGCTCG 1380  
QY 1381 GAAAGAGCTTCAATGAGAGATATGCGGGGTGTGCGACCCCGTGAAGGAGAGGAT 1440  
Db 1381 GAAAGAGCTTCAATGAGAGATATGCGGGGTGTGCGACCCCGTGAAGGAGAGGAT 1440  
QY 1441 GCGACACCACTCCCGTGAAGTCCAGAGAGCCAGAGAGTCTGGGGAGCCCATGGGTC 1500  
Db 1441 GCGACACCACTCCCGTGAAGTCCAGAGAGCCAGAGAGTCTGGGGAGCCCATGGGTC 1500  
QY 1501 CATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGAGCCCGGGGTTCCG 1560  
Db 1501 CATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGAGCCCGGGGTTCCG 1560  
QY 1561 AGGAAAGCCCTCTGAGAGAGAAACCTCCAGCAGCAGGCTGAGAGCTGAGTGTCCAA 1620  
Db 1561 AGGAAAGCCCTCTGAGAGAGAAACCTCCAGCAGCAGGCTGAGAGCTGAGTGTCCAA 1620  
QY 1621 TCCGGGTTGTGAAACCAAGACTTCCCTGAGAGAACAGAGTGCACAGTGAAGGCC 1680  
Db 1621 TCCGGGTTGTGAAACCAAGACTTCCCTGAGAGAACAGAGTGCACAGTGAAGGCC 1680  
QY 1681 AAAGCTGAAGGCTTCTCCCGCAGCCTTCCCGCCCGGGTGTGATGCTGAGAGAG 1740  
Db 1681 AAAGCTGAAGGCTTCTCCCGCAGCCTTCCCGCCCGGGTGTGATGCTGAGAGAG 1740  
QY 1741 TGGCCCTGTGTGAGATGCGGAGAGAGAGTGGCTCATGATGCTGTGCTCCGGTGG 1800  
Db 1741 TGGCCCTGTGTGAGATGCGGAGAGAGAGTGGCTCATGATGCTGTGCTCCGGTGG 1800  
QY 1801 AATGTTCAAGAGTGGCGGTGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGCAATGA 1860  
Db 1801 AATGTTCAAGAGTGGCGGTGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGCAATGA 1860  
QY 1861 CCGAGGTGGCTTGTGTGAGAGAGAGAGAGAGAGTGGCTTGGGGGGGCCCTTGAGCTT 1920  
Db 1861 CCGAGGTGGCTTGTGTGAGAGAGAGAGAGAGAGTGGCTTGGGGGGGCCCTTGAGCTT 1920  
QY 1921 GGAACAGATGGAG 1980  
Db 1921 GGAACAGATGGAG 1980  
QY 1981 GCAAGCTGAG 2040  
Db 1981 GCAAGCTGAG 2040  
QY 2041 GACTACCAAGATTTTAAACAGAAATGTTTAAATTTAATTCCATTTATTA 2100  
Db 2041 GACTACCAAGATTTTAAACAGAAATGTTTAAATTTAATTCCATTTATTA 2100  
QY 2101 ATGTTGGCCACAATTAATGATTAATCTTGTCTGTACTTATGATTTTTCACATTTGT 2160  
Db 2101 ATGTTGGCCACAATTAATGATTAATCTTGTCTGTACTTATGATTTTTCACATTTGT 2160  
QY 2161 GAAAGAACTTAAACAGTAAATGTATGTGTGCGAGTGTGTGTGTGTGTGTGTGTGT 2220  
Db 2161 GAAAGAACTTAAACAGTAAATGTATGTGTGTGCGAGTGTGTGTGTGTGTGTGTGT 2220  
QY 2221 AAAAATGTTGTTTAAAGCTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
Db 2221 AAAAATGTTGTTTAAAGCTTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
QY 2281 GTGAGAGCAAG 2340  
Db 2281 GTGAGAGCAAG 2340  
QY 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
Db 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
RESULT 5  
ACN40903





QY 80 GAGGCGACAGAGGCTACAGTGTCTTACACGCCACAGCCCACTCAAGATATGACAGACC 139  
Db 61 GAGGCGACAGAGGCTACAGTGTCTTACACGCCACAGCCCACTCAAGATATGACAGACC 120  
QY 140 ACCGAGGCTATATGAGGCAAAAGCTATGGAACCTATGACAGGCCACTGATGTCACTAT 199  
Db 121 ACCGAGGCTATATGAGGCAAAAGCTATGGAACCTATGACAGGCCACTGATGTCACTAT 180  
QY 200 ACCGAGGCTCAAGCACTGCAACTATATGAGGCAAGCCCTATGCAACTTCTTATGAGCAG 259  
Db 181 ACCGAGGCTCAAGCACTGCAACTATATGAGGCAAGCCCTATGCAACTTCTTATGAGCAG 240  
QY 260 CCTCCCACTGTTATCTATCTCAACTGCCCCCAGAGGCAATACAGCCAGCTGTCCAGGGG 319  
Db 241 CCTCCCACTGTTATCTATCTCAACTGCCCCCAGAGGCAATACAGCCAGCTGTCCAGGGG 300  
QY 320 TATGCACTGTGCTATATGATACCACTGCTACAGTCAACCAAGCCAGGCTCTCTAT 379  
Db 301 TATGCACTGTGCTATATGATACCACTGCTACAGTCAACCAAGCCAGGCTCTCTAT 360  
QY 380 GCACTCACTGTGCTATATGAGCACTGAGCTGTTATCCAGCTATGAGGCAAGCCAGCA 439  
Db 361 GCACTCACTGTGCTATATGAGCACTGAGCTGTTATCCAGCTATGAGGCAAGCCAGCA 420  
QY 440 GCACTCACTGCTTACAGACCCGAGAGTGAACCAAGCCCACTGAGCTAGTCAACTCA 499  
Db 421 GCACTCACTGCTTACAGACCCGAGAGTGAACCAAGCCCACTGAGCTAGTCAACTCA 480  
QY 500 TCTAGCAAGAGGGGTTTACACAGCCAGCTAGATATGACAGAGTAACTACAGTAT 559  
Db 481 TCTAGCAAGAGGGGTTTACACAGCCAGCTAGATATGACAGAGTAACTACAGTAT 540  
QY 560 CCCCAAGTACCTGAGAGCTACCCCACTGAGCCAGTCACTGCACTCTCACTCTCT 619  
Db 541 CCCCAAGTACCTGAGAGCTACCCCACTGAGCCAGTCACTGCACTCTCACTCTCT 600  
QY 620 ACCAGTATTCCTCTACAGAGCCAGCTAGTATGATCAAGAGCTTACTCTCAGAGAAC 679  
Db 601 ACCAGTATTCCTCTACAGAGCCAGCTAGTATGATCAAGAGCTTACTCTCAGAGAAC 660  
QY 680 ACCATATGAGGCAACGAGAGCTATGACAGCAGAGTATGATGATCAACAAAGCACTAT 739  
Db 661 ACCATATGAGGCAACGAGAGCTATGACAGCAGAGTATGATGATCAACAAAGCACTAT 720  
QY 740 GGGCAGAGCTCTCCACTAGTACCCCACTGAGTCTTACAGCCAGCTCTCAAGT 799  
Db 721 GGGCAGAGCTCTCCACTAGTACCCCACTGAGTCTTACAGCCAGCTCTCAAGT 780  
QY 800 CAATATAGCCCAAGAGAGCAGCTACGGGCAAGAGTTCATTCCGACAGAGCAACCCC 859  
Db 781 CAATATAGCCCAAGAGAGCAGCTACGGGCAAGAGTTCATTCCGACAGAGCAACCCC 840  
QY 860 AGTAGCATGGGTGTTTATGAGGCAAGTCTGAGAGATTTTCCGAGCCAGAGAGAACCCG 919  
Db 841 AGTAGCATGGGTGTTTATGAGGCAAGTCTGAGAGATTTTCCGAGCCAGAGAGAACCCG 900  
QY 920 AGCATGATGGCCCTGATTAACCGGGGCAAGGAGAAAGAGGGGATTTGATCTGAGAGCATG 979  
Db 901 AGCATGATGGCCCTGATTAACCGGGGCAAGGAGAAAGAGGGGATTTGATCTGAGAGCATG 960  
QY 980 AGCAGAGTGGGGGGGAGAGAGACGGGTGGAATGGGGCAGGCTGAGAGAGAGGGTGGC 1039  
Db 961 AGCAGAGTGGGGGGGAGAGAGACGGGTGGAATGGGGCAGGCTGAGAGAGAGGGTGGC 1020  
QY 1040 TTCAATTAAGCCTGTGAGACCATATGATGAAGAGACAGAGTCTTATGAGCCCACTGTA 1099  
Db 1021 TTCAATTAAGCCTGTGAGACCATATGATGAAGAGACAGAGTCTTATGAGCCCACTGTA 1080  
QY 1100 GATTCAGATGAAGCTCTGACAAAGTGCATTTTATGACAGAGTAAATGACAGTGTG 1159  
Db 1081 GATTCAGATGAAGCTCTGACAAAGTGCATTTTATGACAGAGTAAATGACAGTGTG 1140

QY 1160 ACTTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTGTTAAGATGAACAAGAGA 1219  
Db 1141 ACTTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTGTTAAGATGAACAAGAGA 1200  
QY 1220 ACTGGGCAACCCATGATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGCCAT 1279  
Db 1201 ACTGGGCAACCCATGATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGCCAT 1260  
QY 1280 GGCAGAGTCTCTAAGAGACCCCACTGACAGAGGCTGCGGTGAATGTTTATGAGG 1339  
Db 1261 GGCAGAGTCTCTAAGAGACCCCACTGACAGAGGCTGCGGTGAATGTTTATGAGG 1320  
QY 1340 AAAGATTTTCAAGGAGCAAACTTAAAGTCTCTTCTGAGAGAGCCCTCAATGAAC 1399  
Db 1321 AAAGATTTTCAAGGAGCAAACTTAAAGTCTCTTCTGAGAGAGCCCTCAATGAAC 1380  
QY 1400 AGTATGCGGGGTGTCTGTCCACCCCGTGAAGGCAAGGCAATCCCACTCCGTGA 1459  
Db 1381 AGTATGCGGGGTGTCTGTCCACCCCGTGAAGGCAAGGCAATCCCACTCCGTGA 1440  
QY 1460 GGTCCAGAGAGGCGCAGAGAGTCTGAGGGGACCATGGGTGCAATGGGAGGCGGTGA 1519  
Db 1441 GGTCCAGAGAGGCGCAGAGAGTCTGAGGGGACCATGGGTGCAATGGGAGGCGGTGA 1500  
QY 1520 GATTAAGAGAGGCTTCTCTCAAGAGGACCCCGGGGTTCCCGAGGAAACCCCTTGA 1579  
Db 1501 GATTAAGAGAGGCTTCTCTCAAGAGGACCCCGGGGTTCCCGAGGAAACCCCTTGA 1560  
QY 1580 GGAAGGCTCAGACCCAGCTGAGAGCTGAGCACTGTCCCAATCCGGGTGTGGAACAG 1639  
Db 1561 GGAAGGCTCAGACCCAGCTGAGAGCTGAGCACTGTCCCAATCCGGGTGTGGAACAG 1620  
QY 1640 AACTTCGCTGAGAGAACAGAGTGAACAGAGTGAAGGCCCAAGGCTGAGAGGCTCTCTC 1699  
Db 1621 AACTTCGCTGAGAGAACAGAGTGAACAGAGTGAAGGCCCAAGGCTGAGAGGCTCTCTC 1680  
QY 1700 CCGCACCTCTTCCGCCCCCGGGGTGTATCTGTGAGAGAGTGGCCCTGTGATGCGG 1759  
Db 1681 CCGCACCTCTTCCGCCCCCGGGGTGTATCTGTGAGAGAGTGGCCCTGTGATGCGG 1740  
QY 1760 GAGAGAGAGTGTGCTCATGATGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGT 1819  
Db 1741 GAGAGAGAGTGTGCTCATGATGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGT 1800  
QY 1820 GGTGAGAGCAGAGTGTGCTTCCGTGTGCGGGGAGTGAACCGAGAGTGGCTTGTGTA 1879  
Db 1801 GGTGAGAGCAGAGTGTGCTTCCGTGTGCGGGGAGTGAACCGAGAGTGGCTTGTGTA 1860  
QY 1880 GGAAGCAGAGTGTGCTTCCGTGTGCGGGGAGTGAACCGAGAGTGGCTTGTGTA 1939  
Db 1861 GGAAGCAGAGTGTGCTTCCGTGTGCGGGGAGTGAACCGAGAGTGGCTTGTGTA 1920  
QY 1940 AGAGAGAGCCTGAGAGACCTGGAATAATGATTAAGGAGAGCAGCTCAGAGGCCAG 1999  
Db 1921 AGAGAGAGCCTGAGAGACCTGGAATAATGATTAAGGAGAGCAGCTCAGAGGCCAG 1980  
QY 2000 GATCGGCCCTTACATGATGAGAGACCCGAGAGCTGATGATGATGATGATGATGAT 2059  
Db 1981 GATCGGCCCTTACATGATGAGAGACCCGAGAGCTGATGATGATGATGATGATGAT 2040  
QY 2060 TAAACAGAGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2119  
Db 2041 TAAACAGAGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2100  
QY 2120 GATTAATCTTGTCTGTACTTATGATTTTCACTTTGTGAAGAACTTAAACAG 2179  
Db 2101 GATTAATCTTGTCTGTACTTATGATTTTCACTTTGTGAAGAACTTAAACAG 2160  
QY 2180 TTAATGATGATGAGTGTGAGATTTTCTTCTCTTCTTAAATTTAATTTAATTTAAT 2239  
Db 2161 TTAATGATGATGAGTGTGAGATTTTCTTCTCTTCTTAAATTTAATTTAATTTAAT 2220  
QY 2240 TTTAACAATGGGAACCCCTTGTGAGATGCTCAGATCATTTGTGAGAACCAAGAGGCC 2299







```

Db      781  CTACAGCAAGCTCCAAAGTCAATATATAGCAAGAGAGAGCTACGGGAGAGAGTTCC 840
Qy      841  ATTCCGACAGAGACACCCGATGACATGAGTGTATATGAGCAGAGATCTGAGAGATTTTC 900
Db      841  ATTCCGACAGAGACACCCGATGACATGAGTGTATATGAGCAGAGATCTGAGAGATTTTC 900
Qy      901  CGGACCGAGAGAGACCGGAGATGAGTGGCCCTGATTAACGGGGGAGGGGAGAGGGG 960
Db      901  CGGACCGAGAGAGACCGGAGATGAGTGGCCCTGATTAACGGGGGAGGGGAGAGGGG 960
Qy      961  ATTGATCGTGAAGGATGAGCAGAGGTGGCGGGAGAGAGACGGGTGGAATGGGCA 1020
Db      961  ATTGATCGTGAAGGATGAGCAGAGGTGGCGGGAGAGAGACGGGTGGAATGGGCA 1020
Qy      1021  CGCTGAGAGAGCGAGTGGCTTCAATTAAGCTGGTGAACCATGATGAGAGACCAATCT 1080
Db      1021  CGCTGAGAGAGCGAGTGGCTTCAATTAAGCTGGTGAACCATGATGAGAGACCAATCT 1080
Qy      1081  TGATCTAGAGCCCACTGTATGATCAGATGAAGACTCTGACACAGTGCATTTATATACA 1140
Db      1081  TGATCTAGAGCCCACTGTATGATCAGATGAAGACTCTGACACAGTGCATTTATATACA 1140
Qy      1141  AGGATTAATGACAGTGTGACTCTAGATGATCTGACAGACTCTTTAAGCAGTGGGCT 1200
Db      1141  AGGATTAATGACAGTGTGACTCTAGATGATCTGACAGACTCTTTAAGCAGTGGGCT 1200
Qy      1201  TGTTAAGATGAACAAGAACTGGGCAACCATGATCTCAATCTACTGAGCAAGAAAC 1260
Db      1201  TGTTAAGATGAACAAGAACTGGGCAACCATGATCTCAATCTACTGAGCAAGAAAC 1260
Qy      1261  AGGAAAGCCCAAGAGGAGATGCAAGTGTCTTAAGAGACCCCACTGCAAGGCTG 1320
Db      1261  AGGAAAGCCCAAGAGGAGATGCAAGTGTCTTAAGAGACCCCACTGCAAGGCTG 1320
Qy      1321  CGTGAATGCTTTGATGGAAGAAATTTCAAGGAGCAAACTTAAGTCTCCCTGCTG 1380
Db      1321  CGTGAATGCTTTGATGGAAGAAATTTCAAGGAGCAAACTTAAGTCTCCCTGCTG 1380
Qy      1381  GAAGAAAGCTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCTGAGGCGAGAGCAT 1440
Db      1381  GAAGAAAGCTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCTGAGGCGAGAGCAT 1440
Qy      1441  GCCACCAACAATCCGTGAGAGATGAGAGGCGCCAGAGAGTCTTGGGGGAGCCCAATGGGCTG 1500
Db      1441  GCCACCAACAATCCGTGAGAGATGAGAGGCGCCAGAGAGTCTTGGGGGAGCCCAATGGGCTG 1500
Qy      1501  CATGGAGGCGCTGAGAGATGAGAGGCTTCCCTCAAGAGAACCCCGGGGTTCCCG 1560
Db      1501  CATGGAGGCGCTGAGAGATGAGAGGCTTCCCTCAAGAGAACCCCGGGGTTCCCG 1560
Qy      1561  AGGAAACCCCTCTGAGAGAGAAACGTCAGACACCGAGCTGAGAGCTGGACAGTCTCCAA 1620
Db      1561  AGGAAACCCCTCTGAGAGAGAAACGTCAGACACCGAGCTGAGAGCTGGACAGTCTCCAA 1620
Qy      1621  TCCGGGTTGTGAACCAAGAACTTCCCTGAGAAACAGATGCAACAGTGTAAAGGCC 1680
Db      1621  TCCGGGTTGTGAACCAAGAACTTCCCTGAGAAACAGATGCAACAGTGTAAAGGCC 1680
Qy      1681  AAAGCCTGAAGGCTTCTCCGCCACCTTTCCGCCCCCGGGTGTGATCTGAGCAGAGG 1740
Db      1681  AAAGCCTGAAGGCTTCTCCGCCACCTTTCCGCCCCCGGGTGTGATCTGAGCAGAGG 1740
Qy      1741  TGGCCCTGTGTGCAATGCGGGAGAGAGAGTGGCTTATGATCTGTGTGCTCCGGTGG 1800
Db      1741  TGGCCCTGTGTGCAATGCGGGAGAGAGAGTGGCTTATGATCTGTGTGCTCCGGTGG 1800
Qy      1801  AATGTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGGCAATGA 1860
Db      1801  AATGTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGGCAATGA 1860
Qy      1861  CCGAGGTGCTTTGTGTGAGAGAACAGAGTGGCTTGGGGGGCCCTGAGACTTTGAT 1920

```

```

Db      1861  CCGAGGTGCTTTGTGTGAGAGAACAGAGTGGCCCTGGGGGGCCCTGACCTTTGAT 1920
Qy      1921  GGAACAGATGGAGAGAAAGAGAGAGAGTGTGAGAGACTGTGAAAAATGATTAAGCGG 1980
Db      1921  GGAACAGATGGAGAGAAAGAGAGAGAGTGTGAGAGACTGTGAAAAATGATTAAGCGG 1980
Qy      1981  GCAACGTCAGAGCGCAGAGATCGGCTTACTAGATGACAGAGACCCGAGAGCTGCAAT 2040
Db      1981  GCAACGTCAGAGCGCAGAGATCGGCTTACTAGATGACAGAGACCCGAGAGCTGCAAT 2040
Qy      2041  GACTACAGATTTATTTTAAACAGAAATGTTTTAATTAATTAATTAATTAATTAATTA 2100
Db      2041  GACTACAGATTTATTTTAAACAGAAATGTTTTAATTAATTAATTAATTAATTAATTA 2100
Qy      2101  ATGTTGGCCACAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
Db      2101  ATGTTGGCCACAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
Qy      2161  GAAGAAACATTAATAACAGTAAATGATGAGTGGAGTGTGTTTCTCTCTCTCT 2220
Db      2161  GAAGAAACATTAATAACAGTAAATGATGAGTGGAGTGTGTTTCTCTCTCTCTCT 2220
Qy      2221  AAAATGTTGTTTAAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCATATCAAT 2280
Db      2221  AAAATGTTGTTTAAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCATATCAAT 2280
Qy      2281  GTGGAACCAAGAGGCGCTTTAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCATATCAAT 2340
Db      2281  GTGGAACCAAGAGGCGCTTTAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCATATCAAT 2340
Qy      2341  TTTTAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2373
Db      2341  TTTTAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2373

RESULT 8
AAS70647
ID AAS70647 standard; cDNA, 2388 BP.
XX
AC AAS70647,
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6451.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT.
XX
DR WPI, 2001-639362/73.
XX
DR P-PDB; ABG06460.
XX
PT New isolated polynucleotide and encoded polypeptide, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 6451; 103bp; English.
XX

```





QY 702 ATGACAGCAGAGTATGTCTCAACAAGCAGCTATGGGCAAGACCTTCCCACTAGTT 761  
 DB 1608 ATGACAGCAGAGTATGTCTCAACAAGCAGCTATGGGCAAGACCTTCCCACTAGTT 1549  
 QY 762 ACCACCCCAACTGATCTCTCAAGCAGCACTCCCACTCAATATATAGCCAAAGAGCAGCA 821  
 DB 1548 ACCACCCCAACTGATCTCTCAAGCAGCACTCCCACTCAATATATAGCCAAAGAGCAGCA 1489  
 QY 822 GCTACGGGCAAGAGTTCAATTCGACAGGACCACTCCAGTACATGGGTGTTATGGGC 881  
 DB 1488 GCTACGGGCAAGAGTTCAATTCGACAGGACCACTCCAGTACATGGGTGTTATGGGC 1429  
 QY 882 AGGAGCTGTGAGATTTTCCGCAACAGAGAGAACCCGAGCATATAGTGGCCCTGATATACC 941  
 DB 1428 AGGAGCTGTGAGATTTTCCGCAACAGAGAGAACCCGAGCATATAGTGGCCCTGATATACC 1369  
 QY 942 GGGGCAAGGAGAGAGGGGAGATTTGATGTGAGGATGAGAGAGTGGGCGGGAGAGAG 1001  
 DB 1368 GGGGCAAGGAGAGAGGGGAGATTTGATGTGAGGATGAGAGAGTGGGCGGGAGAGAG 1309  
 QY 1002 GACGCGGTGAGATGGGAGCGCTGAGAGAGAGGTGCTTCAATTAAGCTGTGAGACCA 1061  
 DB 1308 GACGCGGTGAGATGGGAGCGCTGAGAGAGAGGTGCTTCAATTAAGCTGTGAGACCA 1252  
 QY 1062 TGGATGAAGAGCAAGATCTTGAATCTTAGGCCCACTGTAGATCCAGATGAAGACTTGACA 1121  
 DB 1251 TGGATGAAGAGCAAGATCTTGAATCTTAGGCCCACTGTAGATCCAGATGAAGACTTGACA 1192  
 QY 1122 ACAGTCAATTTTATGATCAAGATTTAAATGACAGTGTGATCTAGATGATCTGGAGACT 1181  
 DB 1191 ACAGTCAATTTTATGATCAAGATTTAAATGACAGTGTGATCTAGATGATCTGGAGACT 1132  
 QY 1182 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCACA 1241  
 DB 1131 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCACA 1072  
 QY 1242 TCTAAGCTGACAGAGAAACAGAAAGCCCAAGGCCATGCAAGTGTCTTATGAAGACC 1301  
 DB 1071 TCTAAGCTGACAGAGAAACAGAAAGCCCAAGGCCATGCAAGTGTCTTATGAAGACC 1012  
 QY 1302 CACCCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGGAAGATTTTCAAGGAGCAAC 1361  
 DB 1011 CACCCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGGAAGATTTTCAAGGAGCAAC 952  
 QY 1362 TTAAGTCTCTCTTGTCTCGAAGAAAGCTTCAATGAACATATGCGGGGTGTCTGCCAC 1421  
 DB 951 TTAAGTCTCTCTTGTCTCGAAGAAAGCTTCAATGAACATATGCGGGGTGTCTGCCAC 892  
 QY 1422 CCGGTGAGGCAAGAGCATGCTCCACACACTCCGTGAGGATCCAGAGGCCCAAGAGTCC 1481  
 DB 891 CCGGTGAGGCAAGAGCATGCTCCACACACTCCGTGAGGATCCAGAGGCCCAAGAGTCC 832  
 QY 1482 CTGGGGGAGCCCATGGGTGCGATGAGGAGCGGTGAGAGAGATAGAGAGGCTTCCCTCCAA 1541  
 DB 831 CTGGGGGAGCCCATGGGTGCGATGAGGAGCGGTGAGAGAGATAGAGAGGCTTCCCTCCAA 772  
 QY 1542 GAGGAGCCCGGGGTTCCCGAGGAAACCTCTGAGGAGGAAAGCTTCAGACCGAGCTG 1601  
 DB 771 GAGGAGCCCGGGGTTCCCGAGGAAACCTCTGAGGAGGAAAGCTTCAGACCGAGCTG 712  
 QY 1602 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAAACGAATCTTCCCTGAGAAACAGAGT 1661  
 DB 711 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAAACGAATCTTCCCTGAGAAACAGAGT 652  
 QY 1662 GCAACCAAGTGTAAAGCCCAAGAGCTGAAGGCTTCCCTCCGCAACCTTCCGCCCCCG 1721  
 DB 651 GCAACCAAGTGTAAAGCCCAAGAGCTGAAGGCTTCCCTCCGCAACCTTCCGCCCCCG 643  
 QY 1722 GTGGTATGTGCAAGAGTGGCCCTGTGCAATGCGGGAGAGAGAGTGGCTGATGAG 1781  
 DB 642 GTGGTATGTGCAAGAGTGGCCCTGTGCAATGCGGGAGAGAGAGTGGCTGATGAG 583

QY 1782 ATGATGTGTGTCCTGGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCC 1841  
 DB 582 ATGATGTGTGTCCTGGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCC 523  
 QY 1842 GTGGTGGCCGGGCAATGACCCGAGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGAGG 1901  
 DB 522 GTGGTGGCCGGGCAATGACCCGAGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGAGG 463  
 QY 1902 GGGCCCTGTGACCTTTGATGAACAGATGGGAGAAAGAGAGAGAGCTGTGAGACCTG 1961  
 DB 462 GGGCCCTGTGACCTTTGATGAACAGATGGGAGAAAGAGAGAGAGCTGTGAGACCTG 403  
 QY 1962 GAAAAATGATTAAGGCGACCAACCTTCAGAGAGCGAGATTCGGCCCTACTATATGAGA 2021  
 DB 402 GAAAAATGATTAAGGCGACCAACCTTCAGAGAGCGAGATTCGGCCCTACTATATGAGA 343  
 QY 2022 GACCCCGAGAGCTGATGATGATCAACAGATTTATTTTAAACAGAAAAATGTTTAAAT 2081  
 DB 342 GACCCCGAGAGCTGATGATGATCAACAGATTTATTTTAAACAGAAAAATGTTTAAAT 283  
 QY 2082 TTATAATTCATATTTATTAATGTTGGCAACAACATTATGATTTCTGTCTGACTTT 2141  
 DB 282 TTATAATTCATATTTATTAATGTTGGCAACAACATTATGATTTCTGTCTGACTTT 223  
 QY 2142 AGTATTTTCAACATTTGTGAGAAACATTTAAACAAAGTTAATGTAAGTGTGGAGTT 2201  
 DB 222 AGTATTTTCAACATTTGTGAGAAACATTTAAACAAAGTTAATGTAAGTGTGGAGTT 163  
 QY 2202 TTTTTCCTCTCTCTTTTAAATAGTTGTTAAGCTTTAACAATGGGAAACCCCTGTG 2261  
 DB 162 TTTTTCCTCTCTCTTTTAAATAGTTGTTAAGCTTTAACAATGGGAAACCCCTGTG 103  
 QY 2262 GAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCTCTTAACTGTAACTATGTCAT 2321  
 DB 102 GAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCTCTTAACTGTAACTATGTCAT 43  
 QY 2322 GGTGTGATGTTTGT 2363  
 DB 42 GGTGTGATGTTTGT 1

RESULT 10  
 AAS62262/c  
 ID AAS62262 standard; cDNA; 2176 BP.  
 XX  
 AC AAS62262;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE cDNA sequence #49 encoding novel human secreted protein.  
 XX  
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 KW immunosuppressive; antineumatic; ss.  
 OS  
 OS Homo sapiens.  
 XX  
 PN MO200177291-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 XX 29-MAR-2001; 2001WO-US010485.  
 XX  
 XX 06-APR-2000; 2000US-0195604P.  
 PR  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulukota K, Graham JR;  
 XX  
 DR MPI; 2002-010900/01.

PT New polynucleotides encoding secreted proteins useful for treating e.g.  
 PT asthma, HIV and Crohn's disease.

PS Claim 1; Page 100; 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
 CC which encode human secreted proteins. The cDNA sequences have been  
 CC derived from a variety of human tissues. The invention also provides a  
 CC method for producing proteins from these polynucleotide sequences. The  
 CC proteins are useful for identifying compounds that modulate their  
 CC activity and production, and the cell is also useful for identifying  
 CC compounds that modulate expression of the polynucleotide sequences  
 CC encoding the secreted proteins. The sequences of the invention are useful  
 CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
 CC immune deficiency disorders (e.g. severe combined immunodeficiency  
 CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
 CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
 CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
 CC the invention are also useful in gene therapy. AAS62214-AAS62838  
 CC represent the cDNA sequences of the invention that encode for novel human  
 CC secreted proteins

XX Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Query Match 89.6%; Score 2141.8; DB 6; Length 2176;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2154; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 33 AAGAGAGAAATGGCGTCCAGGATTAAGTACCTATAGCCAAAGCTGACGCGACAGG 92  
 Db 2171 AAGGTAGAAATAGGGGTCCAGGATTAAGTACCTATAGCCAAAGCTGACGCGACAGG 2112  
 QY 93 GCTACAGTGTTCACCGCCACCACTCAAGATATGACAGACACCCAGGCTATG 152  
 Db 2111 GCTACAGTGTTCACCGCCACCACTCAAGATATGACAGACACCCAGGCTATG 2052  
 QY 153 GGCACAAAGCTATGAGAACTTATGACAGCCCACTGATGCTATACCCAGGCTCA 212  
 Db 2051 GGCACAAAGCTATGAGAACTTATGACAGCCCACTGATGCTATACCCAGGCTCA 1992  
 QY 213 CCACTGCACCTATGGGACAGCCGCTTATGCACTTTATGACAGCCCTCCAGTGT 272  
 Db 1991 CCACTGCACCTATGGGACAGCCGCTTATGCACTTTATGACAGCCCTCCAGTGT 1932  
 QY 273 ATATCTACTCCATGCCCCCAGGACATACAGCCAGCTGTCCAGGGATATGAGACTG 332  
 Db 1931 ATATCTACTCCATGCCCCCAGGACATACAGCCAGCTGTCCAGGGATATGAGACTG 1872  
 QY 333 CTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGACGCTCACTG 392  
 Db 1871 CTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGACGCTCACTG 1812  
 QY 393 CATATGGCACTGACCTGCTTATCCAGCTTATGGGACAGCAGCAGCACTGCACT 452  
 Db 1811 CATATGGCACTGACCTGCTTATCCAGCTTATGGGACAGCAGCAGCACTGCACT 1752  
 QY 453 CAAGACCGGAGATGAGAAACAGCCCACTGATGCACTTCAATCTAGACAGGGG 512  
 Db 1751 CAAGACCGGAGATGAGAAACAGCCCACTGATGCACTTCAATCTAGACAGGGG 1693  
 QY 513 GTTACAAACAGCCAGCTGATATGAGACAGATTAAGTATCCAGGTAAGT 572  
 Db 1692 GTTACAAACAGCCAGCTGATATGAGACAGATTAAGTATCCAGGTAAGT 1633  
 QY 573 GAGAGTACCCATGACAGCAGTCACTGACCTTCATCTTCTTACAGGTAATCT 632  
 Db 1632 GAGAGTACCCATGACAGCAGTCACTGACCTTCATCTTCTTACAGGTAATCT 1573  
 QY 633 CTACACAGCCGATAGTATGATCAAGAGCTTACTCTGACGAGAACTATAGGGGAC 692  
 Db 1572 CTACACAGCCGATAGTATGATCAAGAGCTTACTCTGACGAGAACTATAGGGGAC 1513  
 QY 693 CGAGCAGCTATGACAGCAGATAGTATGATGCAAAAGCAGCTATGGGACAGGCTC 752

Db 1512 CGAGCAGCTATGACAGCAGATAGCTATGTCACAAAGCAGCTATGGGACAGGCTC 1453  
 QY 753 CCACTATGTTACCCACCCCAACTGATCTTACAGCCAGCTCCAGTCAATATGCAAC 812  
 Db 1452 CCACTATGTTACCCACCCCAACTGATCTTACAGCCAGCTCCAGTCAATATGCAAC 1393  
 QY 813 AAGAGCAGCTACGGGAGCAGAGTTCATTCGACAGGACCAACCCAGTACATGAGTGG 872  
 Db 1392 AAGAGCAGCTACGGGAGCAGAGTTCATTCGACAGGACCAACCCAGTACATGAGTGG 1333  
 QY 873 TTTATGGGAGAGTCTGGAGATTTTCGACAGAGAGAACCGGACATGAGTGGC 932  
 Db 1332 TTTATGGGAGAGTCTGGAGATTTTCGACAGAGAGAACCGGACATGAGTGGC 1273  
 QY 933 CTGATTAACCGGGAGCAGGAGAGAGGGGATTTGATCTGTGAGGATGACAGAGTGGC 992  
 Db 1272 CTGATTAACCGGGAGCAGGAGAGAGGGGATTTGATCTGTGAGGATGACAGAGTGGC 1213  
 QY 993 GGGAGAGAGAGCAGCGGTGAGATGGGACGCTGAGAGAGGAGTGGCTCAATTAAGCTG 1052  
 Db 1212 GGGAGAGAGAGCAGCGGTGAGATGGGACGCTGAGAGAGGAGTGGCTCAATTAAGCTG 1153  
 QY 1053 GTGACCCATGATGAGAGACAGATCTTGAATCTAGGCCACCTGATGATGATGAAG 1112  
 Db 1152 GTGACCCATGATGAGAGACAGATCTTGAATCTAGGCCACCTGATGATGATGAAG 1093  
 QY 1113 ACTTGACAAAGTGCATTTATGTCAGAGATTAATGACAGTGTGATCTTGAATGATC 1172  
 Db 1092 ACTTGACAAAGTGCATTTATGTCAGAGATTAATGACAGTGTGATCTTGAATGATC 1033  
 QY 1173 TGGCAGACTCTTTTAAGCAGTGTGGGGTGTTAAGATGAACAAGAACTGGCAACCA 1232  
 Db 1032 TGGCAGACTCTTTTAAGCAGTGTGGGGTGTTAAGATGAACAAGAACTGGCAACCA 973  
 QY 1233 TGATCCACTTAACTGTCAGCAAGAAACAGAAAGCCCAAGCGATGCAAGTGTCT 1292  
 Db 972 TGATCCACTTAACTGTCAGCAAGAAACAGAAAGCCCAAGCGATGCAAGTGTCT 913  
 QY 1293 ATGAAAGCCCACTGTCAGCAAGAGCTGCGGTGAGATGTTTGAATGGAAAGATTTTCA 1352  
 Db 912 ATGAAAGCCCACTGTCAGCAAGAGCTGCGGTGAGATGTTTGAATGGAAAGATTTTCA 853  
 QY 1353 GAGAGCAACTTAAAGTCTCCTTGTGCGAAGAGGCTCCAAATGAACAGTATCGGGGTG 1412  
 Db 852 GAGAGCAACTTAAAGTCTCCTTGTGCGAAGAGGCTCCAAATGAACAGTATCGGGGTG 793  
 QY 1413 GTCTGCCACCCGTGAGGAGCAGAGGATGCAACCACTCCGTGAGGTCCAGAGGCT 1472  
 Db 792 GTCTGCCACCCGTGAGGAGCAGAGGATGCAACCACTCCGTGAGGTCCAGAGGCT 733  
 QY 1473 CAGAGGCTCTGGGAGACCACTGAGTGCATGAGAGGCGGTGAGAGATGAGAGGCT 1532  
 Db 732 CAGAGGCTCTGGGAGACCACTGAGTGCATGAGAGGCGGTGAGAGATGAGAGGCT 673  
 QY 1533 TCCCTCAGAGAGAACCCGAGGGTCCGAGGGAAACCCCTCTGAGAGAGGAAAGTCCAGC 1592  
 Db 672 TCCCTCAGAGAGAACCCGAGGGTCCGAGGGAAACCCCTCTGAGAGAGGAAAGTCCAGC 613  
 QY 1593 ACCGAGCTGAGATCTGAGAGTCCCAATCCGAGTGTGAGAAACAGAACTTGCCTGAG 1652  
 Db 612 ACCGAGCTGAGATCTGAGAGTCCCAATCCGAGTGTGAGAAACAGAACTTGCCTGAG 553  
 QY 1653 GAAAGAGTGCACACAGTATTAAGGCCCAAGCTTGAAGGCTTCTCCGCGCACCTTTTC 1712  
 Db 552 GAAAGAGTGCACACAGTATTAAGGCCCAAGCTTGAAGGCTTCTCCGCGCACCTTTTC 493  
 QY 1713 CGCCCCGGGTGGTATGTTGGCAGAGGATGCGCTGAGGAGTGGGAGAGAGAGTGT 1772  
 Db 492 CGCCCCGGGTGGTATGTTGGCAGAGGATGCGCTGAGGAGTGGGAGAGAGAGTGT 433  
 QY 1773 GCCTCATGATCGTGTGTCTCCGCTGAGATGTTCAAGGTGAGCGTGTGAGACAGAG 1832

Query Match	85.1%;	Score 2033.4;	DB 13;	Length 2177;
Best Local Similarity	96.4%;	Pred. No. 0;		
Matches 2101;	Conservative 0;	Mismatches 76;	Indels 2;	Gaps 2;
CC	mammals. The invention also relates to nucleic acid and polypeptide			
CC	sequences at least 80% identical to the TAT nucleic acids and			
CC	polypeptides; expression vectors and host cells comprising a TAT nucleic			
CC	acid; an antibody specific for a TAT polypeptide; a peptide or organic			
CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a			
CC	TAT polypeptide; and methods and compositions for the treatment or			
CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,			
CC	antibodies, antagonists, binding molecules and compositions are useful			
CC	for diagnosing or treating a cell proliferative disorder associated with			
CC	increased TAT expression, particularly cancers such as breast cancer,			
CC	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder			
CC	cancer, pancreatic cancer, cervical cancer, cancers of the central			
CC	nervous system, melanoma and leukaemia. TAT nucleic acids may further be			
CC	used as hybridisation probes, in chromosome and gene mapping, in			
CC	chromosome identification and in gene therapy. The present sequence			
CC	represents a TAT nucleic acid of the invention			
XX				
XX				
Sequence 2177 BP; 602 A; 539 C; 605 G; 431 T; 0 U; 0 Other;				
Query Match	85.1%;	Score 2033.4;	DB 13;	Length 2177;
Best Local Similarity	96.4%;	Pred. No. 0;		
Matches 2101;	Conservative 0;	Mismatches 76;	Indels 2;	Gaps 2;
Dy	11	GGACGTTGAGGAAACGAGAGGAAAGAGAGAAAATGGCGTCCAGATTACGTAACCTAT	70	
Db	1	GGAGCTTGAGAGAAACGAGAGGAAAGAGAGAAAATGGCATCCATGATTAACGTAACCTAT	60	
Dy	71	AGCCAACTGAGGGGAGGAGGAGGCTAAGTGCTTACACCGCCGAGCCCACTGAAGATAT	130	
Db	61	AGCCAACTGAGGGGAGGAGGAGGCTAAGTGCTTACAAACCCGAGCCCACTGAAGATAT	120	
Dy	131	GCACAGACCAACCCAGGAGATATGGGCAACAAAGCTATGAACTTATGACAGCCCACTGAT	190	
Db	121	GCACAGACCAACCCAGGAGATATGGGCAACAAAGCTATGAACTTATGACAGCCCACTGAT	180	
Dy	191	GTCAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAAGACCGCTTATGCAATTTCT	250	
Db	181	GTCAGCTATACCCAGGCTCAGACCACTGCAATCTATGGGAGACCGCTTATGCAATTTCT	240	
Dy	251	TATGGAAGAGCTCCCACTGGTTTATACATCTCCAACTGCCCCGAGGATACAGCAAGCT	310	
Db	241	TATGGAAGAGCTCCCACTGGTTTATACATCTCCAACTGCCCCGAGGATACAGCAAGCT	300	
Dy	311	GTCAGAGGGATATGGCACTGGTGCTTATGATACCAACCACTGCTACAGTCAACCAACCCAG	370	
Db	301	GTCAGAGGGATATGGCACTGGTGCTTATGATACCAACCACTGCTACAGTCAACCAACCCAG	360	
Dy	371	GCGTCTCATGCAAGCTCAGTCTGCATATATGGAATCAGGCTGCTTATCCAGCTATGGGAG	430	
Db	361	GCGTCTCATGCAAGCTCAGTCTGCATATATGGAATCAGGCTGCTTATCCAGCTATGGGAG	420	
Dy	431	CAGGCAGAGCACTGCACTGACCAAGACCGAGATGGAAAACAAAGCCCACTGAGACTAGT	490	
Db	421	CAGGCAGATGACATGCACTGACCAAGACTGAGATGGAAAACAAAGCCCACTGAGACTAGT	480	
Dy	491	CAACTCAATCTAGCAAGGGGGGTTTCAACGAGCCGCTAAGGATATGGACAGAGTAAC	550	
Db	481	CAACTCAATCTAGCAAGGGGGGTTTCAACGAGCCGCTAAGGATATGGACAGAGTAAC	540	
Dy	551	TACAGTTATCCCAAGTACTTGGGAGGTAACCCCATGCAAGCAAGTCACTGCACTCATCC	610	
Db	541	TGAGTTATCCCAAGTACTTGGGAGGTAACCCCATGCAAGCAAGTCACTGCACTCATCC	600	
Dy	611	TACCTCTCTACAGCTATTTCTCTTACACAGCCGACTATGTTATGATCAGAGCAAGTACTCT	670	
Db	601	TACCTCTCTACAGCTATTTCTCTTACACAGCCGACTATGTTATGATCAGAGCAAGTACTCT	660	
Dy	671	CAGCAGAAACCTATGGGCAACGAGAGCTATGGAACAGAGAGTACCTATGGCAACAA	730	
Db	661	CAGCAGAAACCTATGGGCAACGAGAGCTATGGAACAGAGAGTACCTATGGCAACAA	720	
Dy	731	AGCAGCTATGGGAGGAGGCTCCCATGATTTACCCACCCAACTGATCTTACAGCAAA	790	

Db 721 AGCAGTATGGGAGCTGCTCCCACTAGTACCACCCCAACTTGATCTACAGCA 780  
 Qy 791 GCTCCAGTAAATATAGCCAAAGAGAGAGCTAGGGAGAGAGATTTCATCCAG 850  
 Db 781 GCTCCAGTAAATATAGCCAAAGAGAGAGCTAGGGAGAGAGATTTCATCCAG 840  
 Qy 851 GACCAACCCAGTACATGAGTGTATATGAGCAGAGTCTGAGAGATTTTCCGACAGCA 910  
 Db 841 GACCAACCCAGTACATGAGTGTATATGAGCAGAGTCTGAGAGATTTTCCGACAGCA 900  
 Qy 911 GAGAACCGGAGCATGATGAGTGGCTGTATACCGGGGAGAGAGAGAGAGATTTGATCGT 970  
 Db 901 GAGAACCGGAGCATGATGAGTGGCTGTATACCGGGGAGAGAGAGAGAGATTTGATCGT 960  
 Qy 971 GAGAGCATGAGCAGAGTGGGGGGGAGAGAGAGCGCGTGGAAATGGGACCGCTGGAGAG 1030  
 Db 961 GAGAGCATGAGCAGAGTGGGGGGGAGAGAGAGCGCGTGGAAATGGGACCGCTGGAGAG 1020  
 Qy 1031 CGAGGTGGCTTCAATAGCCCTGGTGAACCATGGATGAAGACAGATCTTGATCTAGGC 1090  
 Db 1021 CAAGTTGGCTTCAATAGCCCTGGTGAACCATGGATGAAGACAGATCTTGATCTAGGC 1080  
 Qy 1091 CCACTGTAGATCCAGATGAAGACTGTGAACAAGTCAATTTATGTACAGAGATTAAT 1150  
 Db 1081 CCACTGTAGATCCAGATGAAGACTGTGAACAAGTCAATTTATGTACAGAGATTAAT 1140  
 Qy 1151 GACAGTGTACTATAGTATCTGGCAGACTCTTTTAAAGCAGTGTGGGTTGTTAAGATG 1210  
 Db 1141 GACAGTGTACTATAGTATCTGGTGAATCTGTGAGTCTTTTAAAGCAGTGTGGGTTGTTAAGATG 1200  
 Qy 1211 AACAGAGAACTGGGCAACCATGATCAGATCTACCTGGAACAAGAAACAGGAAAGCC 1270  
 Db 1201 AACAGAGAACTGGGCAACCATGATCAGATCTACCTGGAACAAGAAACAGGAAAGCC 1260  
 Qy 1271 AAAGCGCATGCCAGTGTCTTATGAAGACCACTGCAAGCTGCGGTGGAATG 1330  
 Db 1261 AAAGGTATGCTCAGATGTCTGTGAAGACTCACTCTGCAAAAGTCTGCGGTGGAATG 1320  
 Qy 1331 TTTGATGGGAAATTTTCAAGGAGCAAACTTAAATCTCTCTGCTCGGAAGAGCC 1390  
 Db 1321 TTTGATGGGAAATTTTCAAGGAGCAAACTTAAATCTCTCTGCTCGGAAGAGCC 1380  
 Qy 1391 CCAATGAACATGATCCGAGGAGTCTGCAACCCCGTAGGGCAGAGGATGCCACCA 1450  
 Db 1381 CCAATGAACATGATCCGAGGAGTCTGCAACCCCGTAGGGCAGAGGATGCCACCA 1440  
 Qy 1451 CTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCTTGGGAGACCATGGTGCATGGAGGC 1510  
 Db 1441 CTCTGCGGAGGTCCAGAGGCGCCAGAGAGTCTTGGGAGACCATGGTGCATGGAGGC 1500  
 Qy 1511 CGTGAGAGATGAGAGAGTCTTCTTCAAGAGAGACCCGAGGTTTCCGAGGAGACCC 1570  
 Db 1501 CGTGAGAGATGAGAGAGTCTTCTTCAAGAGAGACCCGAGGTTTCCGAGGAGACCC 1560  
 Qy 1571 TCTGAGAGAGAAAGTCCAGACCGAGCTGGAAGTGGAGATGCCAATCCGAGTGT 1630  
 Db 1561 TCTGAGAGAGAAAGTCCAGACCGAGCTGGAAGTGGAGATGCCAATCCGAGTGT 1620  
 Qy 1631 GGAACCAAGAACTTCCGCTGAGAAAGAGTGAACAGTGTAAAGCCCAAAAGCTGAA 1690  
 Db 1621 GGAACCAAGAACTTCCGCTGAGAAAGAGTGAACAGTGTAAAGCTGAA 1680  
 Qy 1691 GGGTTCTCCCGCAACCTTTCCGCGCCCGGGTGTATGCTGGCAGAGTGGCTGTGT 1750  
 Db 1681 GGGTTCTCCCGCAACCTTTCCGCGCCCGGGTGTATGCTGGCAGAGTGGCTGTGT 1740  
 Qy 1751 GGCATGCGGGGAGAGAGAGTGGCTCATGATGATGCTGGTCCCGGTGGAATTTTCA 1810  
 Db 1741 GGCATGCGGGGAGAGAGAGTGGCTCATGATGATGCTGGTCCCGGTGGAATTTTCA 1800  
 Qy 1811 GGTGGCGGTGAGAGCAGAGTGGCTTCCGTGTGGCCGAGGATGAGACCAAGTGGC 1870  
 Db 1801 GGTGGCGGTGAGAGCAGAGTGGCTTCCGTGTGGC-CGAGGATGAGACCAAGTGGC 1859

Qy 1871 TTTGGTGAAGAGAGAGAGTGGCTTGGGGGGCCCTGACCTTTGATGAGACAGATG 1930  
 Db 1860 TTTGGTGAAGAGAGAGAGTGGCTTGGGGGGCCCTGACCTTTGATGAGACAGATG 1918  
 Qy 1931 GAGAGAGAGAGAGAGAGTGGAGAGCTGGAAGAAATGATTAAGGACAGACCGCTCAG 1990  
 Db 1919 GAGAGAGAGAGAGAGAGTGGAGAGCTGGAAGAAATGATTAAGGACAGACCGCTCAG 1978  
 Qy 1991 GAGCGAGAGATCGGCTTCAATAGTGAACACCCCGAGCGTGCATTTGACTACAG 2050  
 Db 1979 GAGCGAGAGATCGGCTTCAATAGTGAACACCCCGAGCGTGCATTTGACTACAG 2038  
 Qy 2051 TTTATTTTAAACGAAATGTTTAAATTAATTCATATTTATTAATTTGGCCA 2110  
 Db 2039 TTTATTTTAAACGAAATGTTTAAATTAATTCATATTTATTAATTTGGCCA 2098  
 Qy 2111 CAACATTATGATTAATCTGTCTGTACTTATGATTTTCAACATTTGTGAAGAAACAT 2170  
 Db 2099 CAACATTATGATTAATCTGTCTGTACTTATGATTTTCAACATTTGTGAAGAAACAT 2158  
 Qy 2171 TAAACAACTTAATGTGA 2189  
 Db 2159 TAAACAACTTAATGTGA 2177

RESULT 12  
 AB199383  
 ID AB199383 standard; cDNA; 2188 BP.  
 XX  
 AC AB199383;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.  
 XX  
 KM Mouse; Ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 OS Mus musculus.  
 XX  
 PN W0200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001MO-JP004192.  
 XX  
 PR 18-MAY-2000; 2000JP-00145977.  
 XX  
 PA (UNYI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 XX  
 PT Ishikawa K, Arai S, Takahashi Y, Negata T, Ishii Y;  
 PT WPI; 2002-034733/04.  
 DR P-FSDB; ABB57126.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 XX  
 PS Claim 2; Page 794-799; 2690pp; English.  
 XX  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an





QY 2041 GACTACCAAGTATTTTAAACGAAATGTTTAAATTATTAATTCATATTATA 2100  
DB 2044 GAGTACCAAGTATTTTAAACGAAATGTTTAAATTATTAATTCATATTATA 2103  
QY 2101 ATGTGGCCCAACATATGATTTCTGTCTGATCTTATGATTTTTCACCATTTGT 2160  
DB 2104 ATGTGGCCCAACATATGATTTCTGTCTGATCTTATGATTTTTCACCATTTGT 2163  
QY 2161 GAAGAAACATTTAAACAGTTAAAT 2185  
DB 2164 GGAGAAACATTTAAACAGTTAAAT 2188

RESULT 13  
ADA53506  
ID ADA53506 standard; cDNA; 1988 BP.  
XX  
AC ADA53506;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human coding sequence, SEQ ID 1074.  
XX  
XX Cytostatic; Anti-inflammatory; Osteoplastic; Neuroprotective; Nootropic;  
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KM Inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN BP1293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
XX  
DR WPI; 2003-395539/38.  
DR P-PSDB; ADA55145.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 1; SEQ ID NO 1074; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

Query Match 75.7%; Score 1808.4; DB 10; Length 1988;  
Best Local Similarity 92.2%; Pred. No. 0;  
Matches 1987; Conservative 0; Mismatches 1; Indels 168; Gaps 1;

QY 33 AAGAGAGAAATAGCGGTCCACGATTAAGTACCTATAGCAAGCTGACGCGAGAG 92  
DB 1 AAGAGAGAAATAGCGGTCCACGATTAAGTACCTATAGCAAGCTGACGCGAGAG 60  
QY 93 GCTACAGTGTTCACCGCCGACCACTCAAGGATATGCAAGACCAACGCGCATATG 152

DB 61 GCTACAGTGTTCACCGCCGACCACTCAAGGATATGCAAGACCAACGCGCATATG 120  
QY 153 GCGACAAAGCTATGGAACCTATGACAGCCCACTGATGCTATACCGAGGCTCAGA 212  
DB 121 GCGACAAAGCTATGGAACCTATGACAGCCCACTGATGCTATACCGAGGCTCAGA 180  
QY 213 CCAGTGCACCTATGGGAGACCGGCTGATCAACTCTTATGAGACAGGCTCCACTGTT 272  
DB 181 CCAGTGCACCTATGGGAGACCGGCTGATCAACTCTTATGAGACAGGCTCCACTGTT 240  
QY 273 ATACTCTCAACTGCCCCCAGGACATACAGCCAGCTCTTCCAGGGGATATGCACTGGT 332  
DB 241 ATACTCTCAACTGCCCCCAGGACATACAGCCAGCTCTTCCAGGGGATATGCACTGGT 300  
QY 333 CTATATATACCACTGCTTACAGTACACCAACCCAGGCTCTTATGAGCTCAGTCTG 392  
DB 301 CTATATATACCACTGCTTACAGTACACCAACCCAGGCTCTTATGAGCTCAGTCTG 360  
QY 393 CATATGGCACTAGGCTGCTTATCCAGCTTATGGGAGAGCCAGGACCACTGACCTTA 452  
DB 361 CATATGGCACTAGGCTGCTTATCCAGCTTATGGGAGAGCCAGGACCACTGACCTTA 420  
QY 453 CAAAGCCGAGATGGAACAAAGCCACTGAGACTAGTCACTCACTCATTTAGCAAGGG 512  
DB 421 CA-----  
QY 513 GTTACACCAAGCCAGCTAGATATGACAGAGTAACTTACAGTTATCCCAAGTACTG 572  
DB 423 -----  
QY 573 GGAGTACCCCATGACAGCACTGCACTGCACTCATCTTACCTTACAGCTATTCCT 632  
DB 423 -----AGCTATTCCT 432  
QY 633 CTACACAGCCGACTAGTATGATCAGAGCAGTTACTCTACAGCAAACTCTATGGCAAC 652  
DB 433 CTACACAGCCGACTAGTATGATCAGAGCAGTTACTCTACAGCAAACTCTATGGCAAC 492  
QY 693 CGAGCAGCTATGAGCAGCAGAGTATGCTATGCTCAAAAGCAGCTATGGGACAGCTCT 752  
DB 493 CGAGCAGCTATGAGCAGCAGAGTATGCTATGCTCAAAAGCAGCTATGGGACAGCTCT 552  
QY 753 CCAGTATGATACCAACCCCAACTGATCTCTACAGCCCACTCTCAATATATGCAAC 812  
DB 553 CCAGTATGATACCAACCCCAACTGATCTCTACAGCCCACTCTCAATATATGCAAC 612  
QY 813 AGAGCAGCAGCTACGGGAGCAGAGTTCAATCCAGACAGACCAACCCAGTATGATGGT 872  
DB 613 AGAGCAGCAGCTACGGGAGCAGAGTTCAATCCAGACAGACCAACCCAGTATGATGGT 672  
QY 873 TTATATGGGAGAGTGTGAGAGTTTCCGAGCCAGAGAGAAACCGGAGCATATGAGTGGC 932  
DB 673 TTATATGGGAGAGTGTGAGAGTTTCCGAGCCAGAGAGAAACCGGAGCATATGAGTGGC 732  
QY 933 CTGATTAACCGGGCAGGGAGAGAGGGGATTTGATCGTGAAGCATAGAGAGTGGG 992  
DB 733 CTGATTAACCGGGCAGGGAGAGAGGGGATTTGATCGTGAAGCATAGAGAGTGGG 792  
QY 993 GGGAGAGAGAGCGGCTGATATGGGAGCGCTGAGAGAGAGAGTGGCTTCAATAGCTG 1052  
DB 793 GGGAGAGAGAGCGGCTGATATGGGAGCGCTGAGAGAGAGAGTGGCTTCAATAGCTG 852  
QY 1053 GTGGAACCATGATGTAAGAGCAAGATCTGATAGGCCCACTGATATCCAGATGAAG 1112  
DB 853 GTGGAACCATGATGTAAGAGCAAGATCTGATAGGCCCACTGATATCCAGATGAAG 912  
QY 1113 ACTCTGACAAAGTGTCAATTTATGTAAGAGATTAATATGACAGTGTGACTCTGATGATC 1172  
DB 913 ACTCTGACAAAGTGTCAATTTATGTAAGAGATTAATATGACAGTGTGACTCTGATGATC 972  
QY 1173 TGGCAGACTTTTAAAGAGTGTGGGTTGTTAAAGTAAACAAGAGACTGGGCAACCA 1232  
DB 973 TGGCAGACTTTTAAAGAGTGTGGGTTGTTAAAGTAAACAAGAGACTGGGCAACCA 1032

QY	1233	IGATCACATCTGACCTGGACAAGGAAACAGGAAAGCCCAAGAGCGCATGCTC	1232
Db	1033	TGATCCACATCTTACTTGGACAGAGAAACAGAAAGCCCAAGAGCGATCCACAGTCT	1092
QY	1233	ATGAAGACCCACCACCTGCGCAAGCTGCGTGTGATGGTATGGTAAGATTTCAAG	1352
Db	1093	ATGAAGACCCACCACCTGCGCAAGCTGCGTGTGATGGTATGGTAAGATTTTCAAG	1152
QY	1353	GGAGCAAACTTAAAGTCTTCCTTGTCTGGAGAAAGCTTCCTCAATGAACGTATGGGGGTG	1412
Db	1153	GGAGCAAACTTAAAGTCTTCCTTGTCTGGAGAAAGCTTCCTCAATGAACGTATGGGGGTG	1212
QY	1413	GTCTGCAACCCCGTGAAGGCGAGGCGATGCCACACCACTCGGTGGAAGTCCAGAGGCG	1472
Db	1213	GTCTGCAACCCCGTGAAGGCGAGGCGATGCCACACCACTCGGTGGAAGTCCAGAGGCG	1272
QY	1473	CAGAGAGTCTTGGGGGAGCCCATGGGTGCAATGGAGAGCCGTGTGAGAGATAGAGAGCT	1532
Db	1273	CAGAGAGTCTTGGGGGAGCCCATGGGTGCGCATGGAGAGCCGTGTGAGAGATAGAGAGCT	1332
QY	1533	TCCCTCCAAAGAGAACCCCGGGGTTCCAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGC	1592
Db	1333	TCCCTCCAAAGAGAACCCCGGGGTTCCAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGC	1392
QY	1593	ACCGAGCTGGAGACTGGGAGTGTCCCAATCCGGGTTGTGAGAAACAGAACTCGCTGGA	1652
Db	1393	ACCGAGCTGGAGACTGGGAGTGTCCCAATCCGGGTTGTGAGAAACAGAACTCGCTGGA	1452
QY	1653	GAACAGAGTGCACACAGTGTAAAGGCCCAAGCCCTGAAGGCTTCTCCCGCACCTTTC	1712
Db	1453	GAACAGAGTGCACACAGTGTAAAGGCCCAAGCCCTGAAGGCTTCTCCCGCACCTTTC	1512
QY	1713	CGCCCCCGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCAATGTGGGGAGAGAAAGTGT	1772
Db	1513	CGCCCCCGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCAATGTGGGGAGAGAAAGTGT	1572
QY	1773	GCCTCATGGATCTGTGTGTGATCCCGGTGGAAATGTTCAAGAGTGTGGCGTGTGTGAGACAGAG	1832
Db	1573	GCCTCATGGATCTGTGTGTGATCCCGGTGGAAATGTTCAAGAGTGTGGCGTGTGTGAGACAGAG	1632
QY	1833	GTGGCTTCCGTGTGTGGCCCGGGCATGACCGAGGTGTGCTTTGTGTGAGAAACGAGGTG	1892
Db	1633	GTGGCTTCCGTGTGTGGCCCGGGCATGACCGAGGTGTGCTTTGTGTGAGAAACGAGGTG	1692
QY	1893	GCCCTGGGGGGGCCCCCTTGACCTTTGATGTGAACAGATGGAGAGAAAGAGAGAGAGCTGT	1952
Db	1693	GCCCTGGGGGGGCCCCCTTGACCTTTGATGTGAACAGATGGAGAGAAAGAGAGAGAGCTGT	1752
QY	1953	GAGGACTGTGAAAAATGTGATAAAGCGAGCAACCGTCAGAGCGGAGAGATGGCGCTACT	2012
Db	1753	GAGGACTGTGAAAAATGTGATAAAGCGAGCAACCGTCAGAGCGGAGAGATGGCGCTACT	1812
QY	2013	AGATGACAGAGACCCCGCAGAGCTGCACTGACCAAGTATTTTAAACAGAAAT	2072
Db	1813	AGATGACAGAGACCCCGCAGAGCTGCACTGACCAAGTATTTTAAACAGAAAT	1872
QY	2073	GTTTTAAATTTATTAATTCATATTTATTAATGTGGCCCAACATTAATATTCCTTGT	2132
Db	1873	GTTTTAAATTTATTAATTCATATTTATTAATGTGGCCCAACATTAATATTCCTTGT	1932
QY	2133	CTGTACTTAAATTTTCAACATTTGTGAAGAAACATTAAACAAAGTTAAATGAT	2188
Db	1933	CTGTACTTAAATTTTCAACATTTGTGAAGAAACATTAAACAAAGTTAAATGAT	1988
RESULT 14			
ADNR07446			
ID	ADNR07446 standard; cDNA; 2026 BP.		
XX			
AC	ADNR07446;		
DT	04-NOV-2004 (first entry)		

Db 181 TGATGAGCTATACCAAGGCTCAGACCACTGACACTATATGGGAGACCGCTTATGAAAC 240  
Qy 247 TTCTTATGACAGCTCTCCACTGTTTACTACTCCAACTGCCCCCAGGACATACAGCA 306  
Db 241 TTCTTATGACAGCTCTCCACTGTTTACTACTCCAACTGCCCCCAGGACATACAGCA 295  
Qy 307 GCCTGTCAGAGGATATGAGCTGTTGTTATGATACCACTGCTTACAGTCAACACAC 366  
Db 296 ----- 295  
Qy 367 CAGAGCTCTTATGACAGCTGCTGATATGAGCACTCAGCTGCTTATGACAGCTATAGG 426  
Db 296 -----TATTCCTCTTACAGAGCACTAGTTATGATCAGAGCTTATCTC 329  
Qy 427 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 486  
Db 340 TCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 399  
Qy 487 TAGTCAACCTCAATCTAGCAAGAGGAGTTACAACAGCAGCAGCAGCAGCAGCAGCAGC 546  
Db 400 TAGTCAACCTCAATCTAGCAAGAGGAGTTACAACAGCAGCAGCAGCAGCAGCAGCAGC 459  
Qy 547 TAACTACAGTTATCTCCAGGATCTGAGAGCTTACCCAGTCAAGCAGCTCACTGACCTCC 606  
Db 460 TAACTACAGTTATCTCCAGGATCTGAGAGCTTACCCAGTCAAGCAGCTCACTGACCTCC 519  
Qy 607 ATCTTACCTCTTACAGCTATCTCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCTTA 666  
Db 520 ATCTTACCTCTTACAGCTATCTCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCTTA 534  
Qy 667 CTCTCAGCAGAACCTATGAGGCAACGAGCAGCTATGACAGAGAGTATGAGTCA 726  
Db 535 -----CAGCTATGACAGCAGAGTATGAGTCA 564  
Qy 727 ACAAGAGCTATGAGGAGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 786  
Db 565 ACAAGAGCTATGAGGAGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 624  
Qy 787 CCAAGCTCCAACTCAATATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 846  
Db 625 CCAAGCTCCAACTCAATATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 684  
Qy 847 ACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 906  
Db 685 ACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 744  
Qy 907 AGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 966  
Db 745 AGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 804  
Qy 967 TCGTGAAGAGATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1026  
Db 805 TCGTGAAGAGATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 864  
Qy 1027 AGAGCAGAGTGTCTTCAATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1086  
Db 865 AGAGCAGAGTGTCTTCAATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 924  
Qy 1087 AGGCCCACTGTATGATCAAGTGAAGCCTGCAACAGTGAATTTATATGACAGAGATT 1146  
Db 925 AGGCCCACTGTATGATCAAGTGAAGCCTGCAACAGTGAATTTATATGACAGAGATT 984  
Qy 1147 AATATGAGTGTGCTTATGATGATCTGAGCACTTCTTAAAGAGTGTGGGTGTTAA 1206  
Db 985 AATATGAGTGTGCTTATGATGATCTGAGCACTTCTTAAAGAGTGTGGGTGTTAA 1044  
Qy 1207 GATGAACAAGAGAGTGGGCAACCATGATCAATCTTACCTGGAACAAGAAACAAGAA 1266  
Db 1045 GATGAACAAGAGAGTGGGCAACCATGATCAATCTTACCTGGAACAAGAAACAAGAA 1104  
Qy 1267 GCCCAAGAGGAGTCCCAAGTGTCTTATGAAGCCCACTGCAAGAGCTGCGGTGA 1326  
Db 1105 GCCCAAGAGGAGTCCCAAGTGTCTTATGAAGCCCACTGCAAGAGCTGCGGTGA 1164

Qy 1327 ATGTTTGAATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGCGAAGAA 1386  
Db 1165 ATGTTTGAATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGCGAAGAA 1224  
Qy 1387 GCCTCAATGAACAGTATGCGGAGTGTCTGCAACCCCGTGAAGGAGCAAGCAATGCCACC 1446  
Db 1225 GCCTCAATGAACAGTATGCGGAGTGTCTGCAACCCCGTGAAGGAGCAAGCAATGCCACC 1284  
Qy 1447 ACCATCTCGTGAAGTCCAGAGAGCCAGAGAGTCTTGGGGAGCCATGGGTGCAATGGG 1506  
Db 1285 ACCATCTCGTGAAGTCCAGAGAGCCAGAGAGTCTTGGGGAGCCATGGGTGCAATGGG 1344  
Qy 1507 AGGCGTGAAGAGATGAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCGAGAGAA 1566  
Db 1345 AGGCGTGAAGAGATGAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCGAGAGAA 1404  
Qy 1567 CCCCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGACTGCACTGTCCCAATCCGAG 1626  
Db 1405 CCCCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGACTGCACTGTCCCAATCCGAG 1464  
Qy 1627 TTGTGAAAACAGAACTTGGCTGAGAAACAGATGCAACCAAGTGAAGGCCCAAGCC 1686  
Db 1465 TTGTGAAAACAGAACTTGGCTGAGAAACAGATGCAACCAAGTGAAGGCCCAAGCC 1524  
Qy 1687 TGAAGGCTCTCCCGGCAACCCCTTCCGCGCGGGTGTGATCGTGAAGAGTGGGCC 1746  
Db 1525 TGAAGGCTCTCTCCCGCAACCCCTTCCGCGCGGGTGTGATCGTGAAGAGTGGGCC 1584  
Qy 1747 TGTGTCATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGATCGCGGTGAATGTT 1806  
Db 1585 TGTGTCATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGATCGCGGTGAATGTT 1644  
Qy 1807 CAGAGTGGCTCGTGTGAGAGCAAGAGTGGCTTCCGTGTGTCGGGGGCAATGGAACGAGG 1866  
Db 1645 CAGAGTGGCTCGTGTGAGAGCAAGAGTGGCTTCCGTGTGTCGGGGGCAATGGAACGAGG 1704  
Qy 1867 TGGCTTTGATGAGAAAGCAAGAGTGGCTCGTGGGGGAGCCCTTGAACCTTTGATGGAACA 1926  
Db 1705 TGGCTTTGATGAGAAAGCAAGAGTGGCTCGTGGGGGAGCCCTTGAACCTTTGATGGAACA 1764  
Qy 1927 GATGAGAGAGAAAGAGAGAGAGCTGAGAGACCTGGAATAATGATTAAGGCGAGCACCG 1986  
Db 1765 GATGAGAGAGAAAGAGAGAGAGCTGAGAGACCTGGAATAATGATTAAGGCGAGCACCG 1824  
Qy 1987 TCAGAGCCGAGAGATCGGCTCTTATGATGAGAGAACCCCGAGAGCTGCAATGACTAC 2046  
Db 1825 TCAGAGCCGAGAGATCGGCTCTTATGATGAGAGAACCCCGAGAGCTGCAATGACTAC 1884  
Qy 2047 CAGATTTATTTTAAACAGAAATGTTTAAATTTATTAATTCATATTTATTAATGTTG 2106  
Db 1885 CAGATTTATTTTAAACAGAAATGTTTAAATTTATTAATTCATATTTATTAATGTTG 1944  
Qy 2107 GCCCAACAATTAATGATTAATCTTGTCTGTAATTAATTTTCAATTTGTGAAGAA 2166  
Db 1945 GCCCAACAATTAATGATTAATCTTGTCTGTAATTAATTTTCAATTTGTGAAGAA 2004  
Qy 2167 ACATTTAAACCAAGTTAAATGCT 2188  
Db 2005 ACATTTAAACCAAGTTAAATGCT 2026

RESULT 15  
ADP56333 standard; cDNA; 1807 BP.  
ID ADP56333  
XX  
XX ADP56333;  
AC  
XX  
XX 18-NOV-2004 (first entry)  
DT  
XX Human PRO cDNA sequence SEQ ID NO:2309.  
DE  
XX human; PRO; immune related disease; inflammatory immune response;  
KM



QY 1201 TGTAAAGTGAACAAGAACTGGGCAACCATGATGCATCTACTGGAACAAGAAC 1260  
| | | | |  
Db 994 TGTAAAGTGAACAAGAACTGGGCAACCATGATGCATCTACTGGAACAAGAAC 1053  
| | | | |  
QY 1261 AGAAGAGCCCAAGGCGATGCGCACATGTCCTATGAAGACCCACCTGCGCAAGGCTGC 1320  
| | | | |  
Db 1054 AGAAGAGCCCAAGGCGATGCGCACATGTCCTATGAAGACCCACCTGCGCAAGGCTGC 1113  
| | | | |  
QY 1321 CGTGAATGTTTGAATGGGAAAGATTTCAGAGGAGCAACTTAAGTCTCCCTGTGCTG 1380  
| | | | |  
Db 1114 CGTGAATGTTTGAATGGGAAAGATTTCAGAGGAGCAACTTAAGTCTCCCTGTGCTG 1173  
| | | | |  
QY 1381 GAAGAACTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCGTGAAGGCGAGGCGAT 1440  
| | | | |  
Db 1174 GAAGAACTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCGTGAAGGCGAGGCGAT 1233  
| | | | |  
QY 1441 GCCACCACTCTCGTGAAGTCCAGAGGCGCCAGAGGTCTTGGGGAGCCCATGGGTGC 1500  
| | | | |  
Db 1234 GCCACCACTCTCGTGAAGTCCAGAGGCGCCAGAGGTCTTGGGGAGCCCATGGGTGC 1293  
| | | | |  
QY 1501 CATGGAGGCGCGTGAAGAGATTAAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTCCCG 1560  
| | | | |  
Db 1294 CATGGAGGCGCGTGAAGAGATTAAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTCCCG 1353  
| | | | |  
QY 1561 AGGGAACCCCTCTGGAAGAGAAAGTTCAGACCCAGCTGAGAGCTGGCAGTGTCCCA 1620  
| | | | |  
Db 1354 AGGGAACCCCTCTGGAAGAGAAAGTTCAGACCCAGCTGAGAGCTGGCAGTGTCCCA 1413  
| | | | |  
QY 1621 TCCGGGTTGTGAAACCAAGACTTTCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCC 1680  
| | | | |  
Db 1414 TCCGGGTTGTGAAACCAAGACTTTCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCC 1473  
| | | | |  
QY 1681 AAAGCCGGAAGGCTTCTCCCGCAACCTTCCCGCCCGCGGTGTGATGCTGCGAGAG 1740  
| | | | |  
Db 1474 AAAGCCGGAAGGCTTCTCCCGCAACCTTCCCGCCCGCGGTGTGATGCTGCGAGAG 1533  
| | | | |  
QY 1741 TGGCCCTGTGTCATGCGGGAGAGAAAGGTGCTCATGGAATGCTGTGTCCTCGGTG 1800  
| | | | |  
Db 1534 TGGCCCTGTGTCATGCGGGAGAGAAAGGTGCTCATGGAATGCTGTGTCCTCGGTG 1593  
| | | | |  
QY 1801 AATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGCGCGGGCATGGA 1860  
| | | | |  
Db 1594 AATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGCGCGGGCATGGA 1653  
| | | | |  
QY 1861 CCGAGGTGGCTTTTGTGTGAAGAAACAGAGTGGCCCTGGGGGGGCCCTTGACCTTTGAT 1920  
| | | | |  
Db 1654 CCGAGGTGGCTTTTGTGTGAAGAAACAGAGTGGCCCTGGGGGGGCCCTTGACCTTTGAT 1713  
| | | | |  
QY 1921 GGAACAGATGGAGGAAGAAAGAGAGACGTGAGAGACTTGAAAAATGATTAAGGCGA 1980  
| | | | |  
Db 1714 GGAACAGATGGAGGAAGAAAGAGAGACGTGAGAGACTTGAAAAATGATTAAGGCGA 1773  
| | | | |  
QY 1981 GCACCGTCAAGAGCGCAGAGATCGGCCCTACTAG 2014  
| | | | |  
Db 1774 GCACCGTCAAGAGCGCAGAGATCGGCCCTACTAG 1807  
| | | | |

Search completed: February 20, 2005, 05:30:30  
Job time : 1175.11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53 ; Search time 396.352 Seconds  
(Without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017A-1

Perfect score: 2390

Sequence: 1 agggggagcagcgcgtgag.....taaaaaaaaaaaaaaaaaa 2390

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/1/ina/5A COMB. seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5B COMB. seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A COMB. seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B COMB. seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/PTUS COMB. seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/backfile1. seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2369.4	99.1	2371	2 US-08-343-443B-1	Sequence 1, Appl
2	2038.4	85.3	6002	4 US-09-949-016-13696	Sequence 13696, A
3	1556	65.1	1785	4 US-09-949-016-5043	Sequence 5043, Ap
4	1556	65.1	1785	4 US-09-949-016-5044	Sequence 5044, Ap
5	1429.6	59.8	1783	4 US-09-949-016-1954	Sequence 1954, Ap
6	819	34.3	2412	1 US-08-437-027-18	Sequence 18, Appl
7	432.6	18.1	35784	4 US-09-949-016-16785	Sequence 16785, A
8	432.6	18.1	35784	4 US-09-949-016-16786	Sequence 16786, A
9	431	18.0	454	4 US-09-513-899C-1657	Sequence 1657, Ap
10	383.4	16.0	425	2 US-08-343-443B-45	Sequence 45, Appl
11	364.2	15.2	601	4 US-09-949-016-66382	Sequence 66382, A
12	338	14.1	411	4 US-09-621-976-13361	Sequence 13361, A
13	306.2	12.8	333	4 US-09-513-999C-12062	Sequence 12062, A
14	266.4	11.1	954	2 US-08-343-443B-106	Sequence 106, App
15	213	8.9	601	4 US-09-949-016-176641	Sequence 176641, Sequence 176641,
16	213	8.9	601	4 US-09-949-016-176642	Sequence 176642, Sequence 322, App
17	175.2	7.3	1939	4 US-09-919-039-322	Sequence 322, App
18	135.8	5.7	601	4 US-09-949-016-176637	Sequence 176637, Sequence 176637,
19	135.8	5.7	601	4 US-09-949-016-176638	Sequence 176638, Sequence 176638,
20	135.8	5.7	601	4 US-09-949-016-176639	Sequence 176639, Sequence 176639,
21	135.8	5.7	601	4 US-09-949-016-176640	Sequence 176640, Sequence 176640,
22	129.4	5.4	601	4 US-09-949-016-176622	Sequence 176622, Sequence 176622,
23	129.4	5.4	601	4 US-09-949-016-176623	Sequence 176623, Sequence 176623,
24	129.4	5.4	601	4 US-09-949-016-176645	Sequence 176645, Sequence 176645,
25	129.4	5.4	601	4 US-09-949-016-176646	Sequence 176646, Sequence 145, App
26	116.4	4.9	450	3 US-09-370-838-145	Sequence 145, App
27	116.4	4.9	450	4 US-09-854-133-145	Sequence 145, App

28	74	3.1	114793	4 US-10-148-806-3	Sequence 3, Appl
29	70.2	2.9	215	4 US-09-513-899C-31510	Sequence 31510, A
30	67	2.8	152132	4 US-09-949-016-13845	Sequence 13845, A
31	67	2.8	152145	4 US-09-949-016-12371	Sequence 12371, A
32	66.4	2.8	311	4 US-09-313-294A-5928	Sequence 5928, A
33	65.6	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
34	64.6	2.7	601	4 US-09-949-016-176640	Sequence 176640, Sequence 176640,
35	64.6	2.7	601	4 US-09-949-016-176641	Sequence 176641, Sequence 19, Appl
36	60	2.5	268	1 US-08-437-027-19	Sequence 19, Appl
37	59.8	2.5	1505	1 US-07-915-246-1	Sequence 1, Appl
38	59	2.5	1225	4 US-09-976-594-416	Sequence 416, App
39	58.8	2.5	34230	4 US-09-949-016-12052	Sequence 12052, A
40	58.8	2.5	128470	4 US-09-949-016-13765	Sequence 13765, A
41	58	2.4	3238	2 US-08-343-443B-5	Sequence 5, Appl
42	57.8	2.4	12695	4 US-09-949-016-16775	Sequence 16775, A
43	57.2	2.4	1682	4 US-09-220-132-82	Sequence 82, Appl
44	57.2	2.4	1684	4 US-09-919-039-323	Sequence 323, App
45	57	2.4	1926	3 US-09-249-585A-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougas, Gilles  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8363  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:





Db	1991	GATGGCCCTTCTAGATCAGAGACCCCGCAGCTGATTCACACAGTTATTTTTT	2049
Qy	2060	TAAACCGAATAATGTTTTAAATTTAAATTCATATTTAAATATGTTGGCCACACATTAAT	2119
Db	2041	TAAACGAAAAATGTTTTAAATTTAAATTCATTTATTAATGTTGGCCACACATTAAT	2100
Qy	2120	GATTAATTCCTGTCTGTACTTTAGATTTTTACACATTGGAAGAAACATTTAAACAG	2179
Db	2101	GATTAATTCCTGTCTGTACTTTAGATTTTTACACATTTGGAGAAACATTTAAACAG	2160
Qy	2180	TTAAATGGTAGTGTGGAGTTTTTTTTCTCTCTTTTAAATAATGGTTTAAAC	2239
Db	2161	TTAAATGGTAGTGTGGAGTTTTTTTTCTCTCTTTTAAATAATGGTTTAAAC	2220
Qy	2240	TTTAAACATGGGAACCCCTTGTGAGCAGTCACATTCATTTGTGAAGAACCAAGGGCC	2299
Db	2221	TTTAAACATGGGAACCCCTTGTGAGCAGTCACATTAATTTGTGAAGAACCAAGGGCC	2280
Qy	2300	TCTTAACTGTAAACATGTTCAATGGTTGTGATGTTTTTTTTTTTTTAAATAAAATTC	2359
Db	2281	TCTTAACTGTAAACATGTTCAATGGTTGTGATGTTTTTTTTTTTTTAAATAAAATTC	2340
Qy	2360	CAATGTTAATTAATAAAAAAAAAAAAAAAAAA	2390
Db	2341	CAATGTTAATTAATAAAAAAAAAAAAAAAAAA	2371

```

RESULT 2
US-09-949-016-13696
; Sequence 13696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13696
; LENGTH: 6002
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-13696

```

Query Match	85.3%	Score 2038.4;	DB 4;	Length 6002;
Best Local Similarity	96.4%;	Pred. No. 0;		
Matches 2106;	Conservative	0;	Mismatches 76;	Indels 2;
			Gaps	2;

Qy	6	GAGACGGAACCTTTGAGAGAAACAGAGCGAGAGAGAGAAAATGGCGTCCACGATTAACAGTA	65
Db	1996	GATGGGGACGTTTGAGAGAACGAGGAGGAGAGAAAATGGCATCCATGGATTACAGTA	2055
Qy	66	CCTTTAGCGAAGCTGCAGCGCAGGAGCTACAGTGTCTTACACCGCCACGCCACTCAAG	125
Db	2056	CCTTTAGCGAAGCTGCAGCGCAGGAGCTACAGTGTCTTACAAACCCAGGCCACTCAAG	2115
Qy	126	GATATGCACAGACCAACCCAGGCATATGGGCAACAAAGTATGAACTTATGACAGCCCA	185
Db	2116	GATATGCACAGACCAACCCAGGCATATGGGCAACAAAGTATGAACTTATGACAGCCCA	2175
Qy	186	CTGATGTTCAGTTATACCAAGGCTCAGACCACTGCAACCTATGGCGAGACGCGCTATGCA	245
Db	2176	TTGATGTTCAGTTATACCAAGGCTCAGACCACTGCAACTTATGGGCGAGACGCGCTATGCA	2235

QY	246	CTTCTTAATGACAGCCTCCCACTGGTTATTA	CTACTCCAACTGCCCCCAGGCATAACAGCC	305
Db	2236	CTTCTTAATGACAGCCTCCCACTGGTTATTA	CTACTCCAACTGCCCCCAGGCATAACAGCC	2295
QY	306	AGCCGTGTACAGGGGTAATGCACTGGTGTAT	TGATACCAACCACTGGCTACACTGACACCA	365
Db	2236	AGCCGTGTACAGGGGTAATGCACTGGTGTAT	TGATACCAACCACTGGCTACACTGACACCA	2355
QY	366	CCAGGCTCCTTAATGACAGCTGATATG	GCACTGAGCTGCTTATCAGCCTAATG	425
Db	2356	CCAGGCTCCTTAATGACAGCTGATATG	GCACTGAGCTGCTTATCAGCCTAATG	2415
QY	426	GGCAGCAGCCGACAGCCCACTGCACTTAC	AGACCGGCAAGATGAAACAAGCCCACTG	485
Db	2416	GGCAGCAGCCGACAGCCCACTGCACTTAC	AGACCGGCAAGATGAAACAAGCCCACTG	2475
QY	486	CTAGTCAACTCAATCTTAGCA	CAGGGGGTTACAAACCGCCAGCTGATGATG	545
Db	2476	CTAGTCAACTCAATCTTAGCA	CAGGGGGTTACAAACCGCCAGCTGATGATG	2535
QY	546	GTAACCTACAGTTATCCCAAGGTACTGG	AGAGCTACCCCATGACAGCTGCACTGCACTC	605
Db	2536	GTAACCTACAGTTATCCCAAGGTACTGG	AGAGCTACCCCATGACAGCTGCACTGCACTC	2595
QY	606	CATCTCACTCCTCTACAGACTAATCTCT	TACACAGCCGACTAGTATGATCAGAGACTT	665
Db	2596	CATCTCACTCCTCTACAGACTAATCTCT	TACACAGCCGACTAGTATGATCAGAGACTT	2655
QY	666	ACTCTCAGCANAACACTTATGCGCA	CCGAGCAGCTTATGACAGCAGATAGCTATGCTC	725
Db	2656	ACTCTCAGCANAACACTTATGCGCA	CCGAGCAGCTTATGACAGCAGATAGCTATGCTC	2715
QY	726	AACAAGCAGCTATGSGCAGAGCCTCC	CACTAGTTATCCCAACCCCAACTGATCTCTACA	785
Db	2716	AACAAGCAGCTATGSGCAGAGCCTCC	CACTAGTTATCCCAACCCCAACTGATCTCTACA	2775
QY	786	GCCAAGCTCCAGTCAATATATAG	CCACAAGACAGCAGCTACGGGACAGAGTTCAATTC	845
Db	2776	GCCAAGCTCCAGTCAATATATAG	CCACAAGACAGCAGCTACGGGACAGAGTTCAATTC	2835
QY	846	GACAGGACCAACCCCAATGACATGGG	GTATATGCGGACAGAGTCTGAGAGATTTTCCGAC	905
Db	2836	GACAGGACCAACCCCAATGACATGGG	GTATATGCGGACAGAGTCTGAGAGATTTTCCGAC	2895
QY	906	CAGAGGAAACCGAGCAGTGAATG	AGCCCTGTAACCGGGGACAGGAGAAAGGGGGATTTG	965
Db	2896	CAGAGGAAACCGAGCAGTGAATG	AGCCCTGTAACCGGGGACAGGAGAAAGGGGGATTTG	2955
QY	966	ATCTGTGAGGAGTGAAGAGAGTGGG	CGGGGAGAGGACGCGGTGGAATGGGACGCGCTG	1025
Db	2956	ATCTGTGAGGAGTGAAGAGAGTGGG	CGGGGAGAGGACGCGGTGGAATGGGACGCGCTG	3015
QY	1026	GAGAGCCGAGGTGCTTCAATTA	AGCTTGATGACCCCATGATGAAGAACAAGATCTTGATC	1085
Db	3016	GAGAGCCGAGGTGCTTCAATTA	AGCTTGATGACCCCATGATGAAGAACAAGATCTTGATC	3075
QY	1086	TAGGCCCACTGTATGATTCAGATGA	AACTCTGCAACATGTGCAATTTATATGACAAGAT	1145
Db	3076	TAGGCCCACTGTATGATTCAGATGA	AACTCTGCAACATGTGCAATTTATATGACAAGAT	3135
QY	1146	TAAATGACAGTGTACTCTAGATAT	CTGCGACAGCTCTTTAAGCAGTGTGGGGTTGTTA	1205
Db	3136	TAAATGACAGTGTACTCTAGATAT	CTGCGACAGCTCTTTAAGCAGTGTGGGGTTGTTA	3195
QY	1206	AGATGAACAGAGAACTGGGCA	ACCCATGATCCACATCTACCTGACAGAGAAACAGAA	1265
Db	3196	AGATGAACAGAGAACTGGGCA	ACCCATGATCCACATCTACCTGACAGAGAAACAGAA	3255
QY	1266	AGCCCAAAAGGGAAGCCACAGTGTCC	ATATGAAGCCCACTGCAAGGCTGCGGTGG	1325
Db	3256	AGCCCAAAAGGGAAGCCACAGTGTCC	ATATGAAGCCCACTGCAAGGCTGCGGTGG	3315
QY	1326	AATGTTGAATGGGAAGATTTTCA	AGGAGCAAACTTAAGTCTCCCTTGCTCGGAAGA	1385

```

Db      3316  AATGTTTATGGAGAAATTTTCAGAGGAGCAACTTAAGTCTCTGCTCGAGAGA 3375
Qy      1386  AGCCTCCATGAAACAGTATCGGGGTGTCTGCCACCCCTGAGGAGAGGATGCGAC 1445
Db      3376  GGCTCCAGTGAACAGTATCAGAGGTGTATGCCACCCCTGAGGAGAGGATGCGAC 3435
Qy      1446  CACCACTTCCTGAGAGTTCAGAGAGCCAGAGGTCTTGGGGGACCCATGGGTGCAATG 1505
Db      3436  CACCACTTCCTGAGAGTTCAGAGAGCCAGAGGTCTTGGGGGACCCATGGGTGCAATG 3495
Qy      1506  GAGGCGGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCGGGGTTCCTCAAGAGA 1565
Db      3496  GAGGCGGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCGGGGTTCCTCAAGAGA 3555
Qy      1566  ACCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
Db      3556  ACACCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3615
Qy      1626  GTTGTGAGAAACCAAACTTGGCTGAGAAACAGATGCCAACAGTGAAGCCCAAGC 1685
Db      3616  GTTGTGAGAAACCAAACTTGGCTGAGAAACAGATGCCAACAGTGAAGCCCAAGC 3675
Qy      1686  CTGAAGGCTTCTCCGCCACCCCTTCCGCCGCCGGGTGATCTGTGAGAGAGAGTGGC 1745
Db      3676  CTGAAGGCTTCTCCGCCACCCCTTCCGCCGCCGGGTGATCTGTGAGAGAGTGGC 3735
Qy      1746  CTGTGTGATGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Db      3736  CTGTGTGATGCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3795
Qy      1806  TCAGAGGTGCGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1865
Db      3796  TCAGAGGTGCGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3854
Qy      1866  GTGCTTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1925
Db      3855  GTGCTTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3913
Qy      1926  AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1985
Db      3914  CAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3973
Qy      1986  GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2045
Db      3974  GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4033
Qy      2046  CCAGATTTATTTTAAACCAAGAAATGTTTAAATTTATTTATTTATTTATTTATTT 2105
Db      4034  CCAGATTTATTTTAAACCAAGAAATGTTTAAATTTATTTATTTATTTATTTATTT 4093
Qy      2106  GGCCCAACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2165
Db      4094  GGCCCAACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4153
Qy      2166  AACATTAACAAAGTTAAATGTA 2189
Db      4154  AACATTAACAAAGTTAAATGTA 4177

```

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5043
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5043

Query Match      65.1%; Score 1556; DB 4; Length 1785;
Best Local Similarity 89.1%; Pred. No. 0;
Matches 1785; Conservative 0; Mismatches 0; Indels 219; Gaps 1;

Qy      11  GAGCGTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70
Db      1  GAGCGTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy      71  AGCCAGCTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
Db      61  AGCCAGCTGACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy      131  GCACAGCACCCAGGAGATATGAGGCAACAAAGCTATGAACTATGAGAGAGAGAGAG 190
Db      121  GCACAGCACCCAGGAGATATGAGGCAACAAAGCTATGAACTATGAGAGAGAGAGAG 180
Qy      191  GTGAGCTATACCCAGGCTGAGACCACTGCAACTATGAGAGAGAGAGAGAGAGAGAG 250
Db      181  GTGAGCTATACCCAGGCTGAGACCACTGCAACTATGAGAGAGAGAGAGAGAGAGAG 240
Qy      251  TATGAGCAAGCTTCCCACTGCTTATATTAATCAACCACTGCTTATGAGAGAGAGAG 310
Db      241  TATGAGCAAGCTTCCCACTGCTTATATTAATCAACCACTGCTTATGAGAGAGAGAG 300
Qy      311  GTCCAGGGATATGAGCACTGAGTGTATGATACCAACCACTGCTTATGAGAGAGAG 370
Db      301  GTCCAGGGATATGAGCACTGAGTGTATGATACCAACCACTGCTTATGAGAGAGAG 360
Qy      371  GCCTCTTATGACAGCTGATGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
Db      361  GCCTCTTATGACAGCTGATGTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      431  CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 490
Db      421  CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      491  CAACCTCAATCTGAGCAAGGGGGTTCAACCAAGCCAGAGAGAGAGAGAGAGAGAG 550
Db      481  CAACCTCAATCTGAGCAAGGGGGTTCAACCAAGCCAGAGAGAGAGAGAGAGAGAG 540
Qy      551  TACAGTTATCCAGAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 610
Db      541  TACAGTTATCCAGAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      611  TACCTCTTACAGAGATATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 670
Db      601  TACCTCTTACAGAGATATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      671  CAGCAGAACCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730
Db      661  CAGCAGAACCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy      731  AGCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
Db      721  AGCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy      791  GCTCCAGATTAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 850
Db      781  GCTCCAGATTAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 826
Qy      851  GACCAACCCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 910

```

```

RESULT 3
US-09-949-016-5043
; Sequence 5043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

```

[illegible]

DB	1702	GGAGAGAAAGAGGAGGAGCGTGGAGACCTGGAAAAATGATATAAAGGCGAGCAACGCTAG	1761
QY	1991	GAGCGCAGAGATCGGCTCTACTAG	2014
DB	1762	GAGCGCAGAGATCGGCTCTACTAG	1785
<p>RESULT 4</p> <p>US-09-949-016-5044</p> <p>/ Sequence 5044, Application US/09949016</p> <p>/ Patent No. 6812339</p> <p>/ GENERAL INFORMATION:</p> <p>/ APPLICANT: VENTER, J. Craig et al.</p> <p>/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED</p> <p>/ TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF</p> <p>/ FILE REFERENCE: C1001307</p> <p>/ CURRENT APPLICATION NUMBER: US/09/949,016</p> <p>/ CURRENT FILING DATE: 2000-04-14</p> <p>/ PRIOR APPLICATION NUMBER: 60/241,755</p> <p>/ PRIOR FILING DATE: 2000-10-20</p> <p>/ PRIOR APPLICATION NUMBER: 60/237,768</p> <p>/ PRIOR FILING DATE: 2000-10-03</p> <p>/ PRIOR APPLICATION NUMBER: 60/231,498</p> <p>/ PRIOR FILING DATE: 2000-09-08</p> <p>/ NUMBER OF SEQ ID NOS: 207012</p> <p>/ SOFTWARE: FastSeq for Windows Version 4.0</p> <p>/ SEQ ID NO 5044</p> <p>/ LENGTH: 1785</p> <p>/ TYPE: DNA</p> <p>/ ORGANISM: Human</p> <p>US-09-949-016-5044</p>			
QY	11	65.1%; Score 1556; DB 4; Length 1785;	
DB	1	89.1%; Pred. No. 0;	
QY	71	Matches 1785; Conservative 0; Mismatches 0; Indels 219; Gaps 1;	
DB	61		
QY	131		
DB	121		
QY	191		
DB	181		
QY	251		
DB	241		
QY	311		
DB	301		
QY	371		
DB	361		
QY	431		
DB	421		
QY	491		
DB	481		
QY	551		

Db	1402	GGAATCCGAACTTTCGCTGTGAGAACAGAGTGCAACCAAGTGTAAAGGCCCCCAAAGCTTGA	1461
Oy	1691	GGCTTTCCCTCCCGCACACCCTTTCCGGCCCCCGGGGTGGTGTATCGTGGCAGAGAGTGACCCTGGT	1750
Db	1462	GGCTTCTCTCCGCCCAACCTTTCTCCGCCCCCGGGGTGTGATCGTGGCAGAGAGTGGCCCTGGT	1521
Oy	1751	GGCATGCGGGGAGGAAGAAGGTGGCCCTCATYGAATCGTGTGTGCTCCGGTGGAAATGTTCA	1810
Db	1522	GGCATGCGGGGAGGAAGAAGGTGGCCCTCATYGAATCGTGTGTGCTCCGGTGGAAATGTTCA	1581
Oy	1811	GGTGGCCGTGTGTGAGACAAGAGGTGGCTTCGTGTGTGGCCGGGGGATGGAACGAGGTGGC	1870
Db	1582	GGTGGCCGTGTGTGAGACAAGAGGTGGCTTCGTGTGTGGCCGGGGGATGGAACGAGGTGGC	1641
Oy	1871	TTTGTGTGAGGAAGACGAGGTGGCCCTTGGGGGGCCCCCTGACCTTTGATYGAAACAGATG	1930
Db	1642	TTTTGTGTGAGGAAGACGAGGTGGCCCTTGGGGGGCCCCCTGACCTTTGATYGAAACAGATG	1701
Oy	1931	GGAGGAAGAAGAGAGGACGTGTGAGAGACTTGAAAATAAGATTAAGGCAGACACGCTCAG	1990
Db	1702	GGAGGAAGAAGAGAGGACGTGTGAGAGACTTGAAAATAAGATTAAGGCAGACACGCTCAG	1761
Oy	1991	GAGCGACAGATCGGCCCTTACTAG 2014	
Db	1762	GAGCCGACAGATCGGCCCTTACTAG 1785	
 RESULT 5 US-09-949-016-1954 Sequence 1954, Application US/09949016 Patent No. 6812339 GENERAL INFORMATION: APPLICANT: VENTER, J. Craig et al. TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF FILE REFERENCE: CLO01307 CURRENT APPLICATION NUMBER: US/09/949, 016 CURRENT FILING DATE: 2000-04-14 PRIOR APPLICATION NUMBER: 60/241,755 PRIOR FILING DATE: 2000-10-20 PRIOR APPLICATION NUMBER: 60/237,768 PRIOR FILING DATE: 2000-10-03 PRIOR APPLICATION NUMBER: 60/231,498 PRIOR FILING DATE: 2000-09-08 NUMBER OF SEQ ID NOS: 207012 SOFTWARE: FastSeq for Windows Version 4.0 SEQ ID NO 1954 LENGTH: 1783 TYPE: DNA ORGANISM: Human US-09-949-016-1954			
 Query Match            59.8%; Score 1429.6; DB 4; Length 1783; Best Local Similarity 85.8%; Pred.No. 0; Matches 1719; Conservative 0; Mismatches 64; Indels 221; Gaps 3;			
Oy	11	GAACGTTAGAGAAAGAGAGAGAAAGAAAAATGCGCTCCAGCGATTACGATCCTAT 70	
Db	1	GAACGTTAGAGAAAGAGAGAGAAAGAAAAATGCGATCCATGATTAACGACTCTAT 60	
Oy	71	AGCCAAGTCGACGCGACGAGGGCTTACAGTGTCTTACACCGCCCACTCAAGANAT 130	
Db	61	AGCCAAGTCGACGCGACGAGGGCTTACAGTGTCTTACACACCCAGCCCACTCAAGANAT 120	
Oy	131	GCACGACACACCCAGGCAATATGGGCAACAAAGCTATGAACTTATGAGACAGCCACTGAT 190	
Db	121	GCACGACACACCCAGGCAATATGGGCAACAAAGCTATGAACTTATGAGACAGCCACTGAT 180	
Oy	191	GTCAAGCTATACCCAGGCTCAGACCACTGCAACTTATATGGCAGACCGCTTATGCACTTCT 250	
Db	181	GTCAAGCTATACCCAGGCTCAGACCACTGCAACTTATATGGCAGACCGCTTATGCACTTCT 240	
Oy	251	TATGAGACAGCTTCCACTGTGTTATACTACTCCACTGCCCCCGGCAATACAGCCAGCT 310	

```

RESULT 5
US-09-949-016-1954
; Sequence 1954, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0010107
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1954
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1954

Query Match          59.8%; Score 1429.6; DB 4; Length 1783;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1719; Conservative 0; Mismatches 64; Indels 221; Gaps 3;

QY      11  GCACGTTGAGAGAACGAGAGGAGAGAGAAATGGCGTCCACGGATTACAGTACCTTAT 70
DB      1  GAGCGTTAGAGAAACGAGAGGAGAGAGAAATATGGCATTCATGATTACAGTACCTAT 60

QY      71  AGCCAGACTGCAGAGCCAGACGAGGCTCACTGCTTACACGCCGCCAGCCACTCAAGATAT 130
DB      61  AGCCAGAGTGCAGCCAGACGAGGCTCACTGCTTACACACCCAGCCACTCAAGATAT 120

QY      131  GCACGAGCACCCACGAGGCAATATGGGCAACAAAGCTATGGAACCTATGACAGGCCCACTGAT 190
DB      121  GCACGAGCACCCACGAGGCAATATGGGCAACAAAGCTATGGAACCTATGACAGGCCCACTTGT 180

QY      191  GTCAAGCTATACCCAGAGCTCAGACCACTGCACCTATGGGAGAGCCGCTATGCACTTCT 250
DB      181  GTCAAGCTATACCCAGAGCTCAGACCACTGCACCTATGGGAGAGCCGCTATGCACTTCT 240

QY      251  TATGAGACAGCTCCCACTGCTTATACTCACTGCCACCTGCCCCCGAGGCAATACAGCCAGCT 310

```

Db 241 TATGAGCAGCCTCCCACTGGTTATACACTCCACCTCCCCCGAGGATACAGCAAGCCT 300  
 Qy 311 GTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCAACCAAGCCT 370  
 Db 301 GTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCAACCAAGCCT 360  
 Qy 371 GCGCTCTATGAGTCACTGCTGATATGGCACTGAGCTGCTTATCCAGCTATGGGAG 430  
 Db 361 GCGCTCTATGAGTCACTGCTGATATGGCACTGAGCTGCTTATCCAGCTATGGGAG 420  
 Qy 431 CAGCAGCAGCAGTCACTGACCTACAGACCGGAGGATGAAACAGCCCACTGAGACTAGT 490  
 Db 421 CAGCAGTACCACTGACCTACAGACCGGAGGATGAAACAGCCCACTGAGACTAGT 480  
 Qy 491 CAACCTCATATCTAGCAAGAGGGGTTTAAACAGCCCACTGAGTATGACAGAGTAA 550  
 Db 481 CAACCTCATATCTAGCAAGAGGGGTTTAAACAGCCCACTGAGTATGACAGAGTAA 540  
 Qy 551 TACAGTATGCCAGTACCTGGGAGGTTACCCCACTGACAGCTCACTGACCTCCATCC 610  
 Db 541 TGCAGTATGCCAGTACCTGGGAGGTTACCCCACTGACAGCTCACTGACCTCCATCC 600  
 Qy 611 TACCTCTCTACAGCTATCTCTCTACACAGCCGACTAGTATGATCAGAGCACTTACT 670  
 Db 601 TACCTCTCTACAGCTATCTCTCTACACAGCCGACTAGTATGATCAGAGCACTTACT 660  
 Qy 671 CAGCAGAACCTATGAGGCAACCGAGCACTATGACAGCAGAGTATGCTTAA 730  
 Db 661 CAGCAGAACCTATGAGGCAACCGAGCACTATGACAGCAGAGTATGCTTAA 720  
 Qy 731 AGAGCTATGAGGCAAGCTCCCACTAGTTACCCCACTGAGTATCTTACAGCA 790  
 Db 721 AGAGCTATGAGGCAAGCTCCCACTAGTTACCCCACTGAGTATCTTACAGCA 780  
 Qy 791 GCTCAAGTATATATAGCAAGAGCAAGCACTGAGGCAAGCACTTATCCAGCAG 850  
 Db 781 GCTCAAGTATATATAGCAAGAGCAAGCACTGAGGCAAGCACTTATCCAGCAG 826  
 Qy 851 GACCAACCCAGTATGAGTGGTGTATGAGGCAAGAGTCTGAGAGATTTTCCGACAG 910  
 Db 827 ----- 826  
 Qy 911 GAGAACCGAGCATGATGAGCCCTGATTAACCGGAGCAGGAGGAGATTGATCCT 970  
 Db 827 ----- 826  
 Qy 971 GAGAGCATGAGCAGAGGTGGGCGGAGAGAGACGCGGTGGAATGGCAGGCTGAGAG 1030  
 Db 827 ----- 826  
 Qy 1031 CGAGGTGGCTTCAATAGACCTGTGTGACCCATGATGAGAGCAGATCTTGTATGAGC 1090  
 Db 827 -----GACCATGATTAAGGACCAAGATCTTGTATGAGC 861  
 Qy 1091 CCACCTGTATATCCAGTGAAGACTCTGACCAAGTCAATTTATGTAAGAGATTAAAT 1150  
 Db 862 CCACCTGTATATCCAGTGAAGACTCTGACCAAGTCAATTTATGTAAGAGATTAAAT 921  
 Qy 1151 GACAGGTGATCTAGATGATCTGGGAGACTCTTAAAGCAGTGTGGGTTGTTAAAGT 1210  
 Db 922 GACAGGTGATCTAGATGATCTGGGAGACTCTTAAAGCAGTGTGGGTTGTTAAAGT 981  
 Qy 1211 AACAGAGAACTGGGCAACCATGATCCATCTTACTGAGCAGAGAAAGAGAAAGCC 1270  
 Db 982 AACAGAGAACTGGGCAACCATGATCCATCTTACTGAGCAGAGAAAGAGAAAGCC 1041  
 Qy 1271 AAAGGCAATGACAGAGTCTTATGAAAGCCCACTGCAAGGCTGCCGTGGAATGG 1330  
 Db 1042 AAAGGCAATGACAGAGTCTTATGAAAGCCCACTGCAAGGCTGCCGTGGAATGG 1101  
 Qy 1331 TTTGATGGAAAGATTTTCAAGGGAGCAATTAATCTCCCTGTGCGAAGAGACCT 1390

Db 1102 TTTGATGGAAAGATTTTCAAGGGAGCAAACTTAAGTCTCTCTGTCGGAAGAGCCT 1161  
 Qy 1391 CCAATGAACATATGCGGGGTGTCTTCCACCCCTGTAAGGCAAGGCAATGCCACCA 1450  
 Db 1162 CCAATGAACATATGCGGGGTGTCTTCCACCCCTGTAAGGCAAGGCAATGCCACCA 1221  
 Qy 1451 CTCGCTGGAAGTCCAGAGGCCCCAGAGAGTCTTGGGGAACCAATGCTGCAATGGAGGC 1510  
 Db 1222 CTCGCTGGAAGTCCAGAGGCCCCAGAGAGTCTTGGGGAACCAATGCTGCAATGGAGGC 1281  
 Qy 1511 CGTGAAGATATGAGAGGCTTCCCTCAAGAGAACCCCGGGTTCCGAGGGAACCC 1570  
 Db 1282 CGTGAAGATATGAGAGGCTTCCCTCAAGAGAACCCCGGGTTCCGAGGGAACCC 1341  
 Qy 1571 TCTGAAGAGAAACGTCACAGCACCGAGCTGGAAGTCTGCAAGTCTCCATCCGGTTGT 1630  
 Db 1342 TCTGAAGAGAAACGTCACAGCACCGAGCTGGAAGTCTGCAAGTCTCCATCCGGTTGT 1401  
 Qy 1631 GGAACCAAGAACTTCCGCTGGAAGACAGAGTGCACACAGGTAAAGCCCAAGCCTGAA 1690  
 Db 1402 GGAACCAAGAACTTCCGCTGGAAGACAGAGTGCACACAGGTAAAGCCTGAA 1461  
 Qy 1691 GCGTCTCTCCCGCACCTTTCCGCCCCCGGGTGTGATCGTGAAGAGTGGCCCTGT 1750  
 Db 1462 GCGTCTCTCCCGCACCTTTCCGCCCCCGGGTGTGATCGTGAAGAGTGGCCCTGT 1521  
 Qy 1751 GCGATGCGGGAGAGAGAGTGGCTCATGATGATGATGATGATGATGATGATGATGAT 1810  
 Db 1522 GCGATGCGGGAGAGAGAGTGGCTCATGATGATGATGATGATGATGATGATGATGAT 1581  
 Qy 1811 GGTGGCCGTGTGGAAGACAGAGTGGCTTCCGTGTGGCGGGAGCAATGACCGAGGTGGC 1870  
 Db 1582 GGTGGCCGTGTGGAAGACAGAGTGGCTTCCGTGTGGCGGGAGCAATGACCGAGGTGGC 1640  
 Qy 1871 TTTGTGGAAGAAAGCAGAGTGGCTTGGGGGCCCCCTGAGACTTTGATGAAACAGATG 1930  
 Db 1641 TTTGTGGAAGAAAGCAGAGTGGCTTGGGGGCCCCCTGAGACTTTGATGAAACAGATG 1699  
 Qy 1931 GAGAGAAAGAGAGAGAGCTGAGAGACTTGAAGAAATGATTAAGGCGAGCAGCTGAG 1990  
 Db 1700 GAGAGAAAGAGAGAGAGCTGAGAGACTTGAAGAAATGATTAAGGCGAGCAGCTGAG 1759  
 Qy 1991 GAGGCGAGAGATGGCCCTACTAG 2014  
 Db 1760 GAGGCGAGAGATGGCCCTACTAG 1783

RESULT 6  
 US-08-437-027-18  
 ? Sequence 18, Application US/08437027  
 ? Patent No. 5670317  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Landany, Marc  
 ? APPLICANT: Gerald, William  
 ? TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
 ? NUMBER OF SEQUENCES: 21  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Cooper & Dunham LLP  
 ? STREET: 1185 Avenue of the Americas  
 ? CITY: New York  
 ? STATE: New York  
 ? COUNTRY: U.S.A.  
 ? ZIP: 10036  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/437,027  
 ? FILING DATE:  
 ? CLASSIFICATION: 536







121 TCAAGATATGCA CAGACCA CCGAGGATATGAGCA CAAAGCTATGAACTATGACA 180  
121 TCAAGATATGCA CAGACCA CCGAGGATATGAGCA CAAAGCTATGAACTATGACA 180  
181 GCCCATGATGATGAGCTATACCAAGGCTGAGACCACTGCAACCTATGAGGAGACCCGCTTA 240  
181 GCCCATGATGATGAGCTATACCAAGGCTGAGACCACTGCAACCTATGAGGAGACCCGCTTA 240  
241 TGCAATCTTTATGAGACGCTCCCACTGTTATATCTACCTCAACCTGCCCCCGGAGATA 300  
241 TGCAATCTTTATGAGACGCTCCCACTGTTATATCTACCTCAACCTGCCCCCGGAGATA 300  
301 CAGCAGACCTGTCAGAGGATATGAGCACTGAGCTTATGATACCACTGCTACAGTAC 360  
301 CAGCAGACCTGTCAGAGGATATGAGCACTGAGCTTATGATACCACTGCTACAGTAC 360  
301 CAGCAGACCTGTCAGAGGATATGAGCACTGAGCTTATGATACCACTGCTACAGTAC 360  
361 CACCA CCGAGGCTCTATGAGCACTGAGCTTATGATACCACTGCTACAGTAC 420  
361 CACCA CCGAGGCTCTATGAGCACTGAGCTTATGATACCACTGCTACAGTAC 420  
421 CTATGAGGACAGGACGAGCACTGAGCTTATGATACCACTGCTACAGTAC 480  
421 CTATGAGGACAGGACGAGCACTGAGCTTATGATACCACTGCTACAGTAC 480  
481 TGAGACTAGTCAACCTCAATCTAGCA CAGAGGAGTTTAA CCAAGCCAGCTAGAGTATGG 540  
481 TGAGACTAGTCAACCTCAATCTAGCA CAGAGGAGTTTAA CCAAGCCAGCTAGAGTATGG 540  
541 ACAGAGTAACTA CAGTTATCTCCAGAGTACCTGAGAGCTACCCCACTGAGCACTGAC 600  
541 ACAGAGTAACTA CAGTTATCTCCAGAGTACCTGAGAGCTACCCCACTGAGCACTGAC 600  
601 ACCTCCATCTTACCTCTACAGCTATCTCTACCA CAGCCCACTGATATGATACAG 660  
601 ACCTCCATCTTACCTCTACAGCTATCTCTACCA CAGCCCACTGATATGATACAG 660  
661 CAGTTACTCTCAGAGAA CACTATGAGGCA CCGAGAGCTATGAGCA CAGAGTACGTA 720  
661 CAGTTACTCTCAGAGAA CACTATGAGGCA CCGAGAGCTATGAGCA CAGAGTACGTA 720  
721 TGCTCAACAAAGCACTATGAGGCA CAGCTCCCACTGATACCACTGATAC 780  
721 TGCTCAACAAAGCACTATGAGGCA CAGCTCCCACTGATACCACTGATAC 780  
781 CTACAGCAAGCTCAAGTCAATATATGACCA CAGAGAGCACTGAGGAGAGAGATTC 840  
781 CTACAGCAAGCTCAAGTCAATATATGACCA CAGAGAGCACTGAGGAGAGAGATTC 840  
841 ATTCCGACAGGACCA CCGAGTACAGTGGTGTATGAGGAGAGATCTGAGAGATTTTC 900  
841 ATTCCGACAGGACCA CCGAGTACAGTGGTGTATGAGGAGAGATCTGAGAGATTTTC 900  
901 CCGACAGGAGAGAA CCGAGAGCACTGAGTGGCTCTGATAC CCGAGGAGAGAGAGGAG 960  
901 CCGACAGGAGAGAA CCGAGAGCACTGAGTGGCTCTGATAC CCGAGGAGAGAGAGGAG 960  
961 ATTGATCGTGAAGGCACTGAGCAAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
961 ATTGATCGTGAAGGCACTGAGCAAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
1021 CGCTGAG 1080  
1021 CGCTGAG 1080  
1081 TGATCTAGGCGCCACTGATGATCAAGTGAAGCTGAGCA CAGTCAATTTATATGACA 1140  
1081 TGATCTAGGCGCCACTGATGATCAAGTGAAGCTGAGCA CAGTCAATTTATATGACA 1140  
1141 AGAATTTAAATGAGAGTGTGATCTAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1141 AGAATTTAAATGAGAGTGTGATCTAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1201 TGTTAAGATGAACAAGAGAACTGGGCAACCATGATCACTTACCTGAGCAAGAGAAAC 1260

1201 TGTTAAGATGAACAAGAGAACTGGGCAACCATGATCACTTACCTGAGCAAGAGAAAC 1260  
1261 AGGAAAGCCCAAGAGAGATGCCAGATGCTATGAAAGCCCACTGACCAAGGCTGC 1320  
1261 AGGAAAGCCCAAGAGAGATGCCAGATGCTATGAAAGCCCACTGACCAAGGCTGC 1320  
1321 CGTGAAGTGGTTGATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTCG 1380  
1321 CGTGAAGTGGTTGATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTCG 1380  
1381 GAAAGAGCTCCAAATGAACAGTATGAGGAGTGTGCA CCGCTGAGAGGAGAGAGAT 1440  
1381 GAAAGAGCTCCAAATGAACAGTATGAGGAGTGTGCA CCGCTGAGAGGAGAGAGAT 1440  
1441 GGCACCA CCACTCCGAGAGATCCAGAGAGGCTCCAGAGAGTCTGAGGAGAGAGAGAGAT 1500  
1441 GGCACCA CCACTCCGAGAGATCCAGAGAGGCTCCAGAGAGTCTGAGGAGAGAGAGAGAT 1500  
1501 CATGAGAGGCGGAGAGAGATGAGAGAGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1501 CATGAGAGGCGGAGAGAGATGAGAGAGGCTCCCAAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1561 AGGAAAGCCCTCTGAG 1620  
1561 AGGAAAGCCCTCTGAG 1620  
1621 TCGAGGTTGAGAAACCAAGACTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1621 TCGAGGTTGAGAAACCAAGACTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1681 AAGAGCTGAAGGCTTCTCCCGCA CCGCTTCCGAGGAGAGAGAGAGAGAGAGAGAGAG 1740  
1681 AAGAGCTGAAGGCTTCTCCCGCA CCGCTTCCGAGGAGAGAGAGAGAGAGAGAGAGAG 1740  
1741 TGAGCCTGAGGAGATGAG 1800  
1741 TGAGCCTGAGGAGATGAG 1800  
1801 AATGTTCAAGAGTGGCTGAG 1860  
1801 AATGTTCAAGAGTGGCTGAG 1860  
1861 CCGAGGAGGCTTGTGAG 1920  
1861 CCGAGGAGGCTTGTGAG 1920  
1861 CCGAGGAGGCTTGTGAG 1920  
1861 CCGAGGAGGCTTGTGAG 1920  
1921 GGAACAGATGAG 1980  
1921 GGAACAGATGAG 1980  
1981 GCAAGGAG 2040  
1981 GCAAGGAG 2040  
1981 GCAAGGAG 2040  
1981 GCAAGGAG 2040  
2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTATTTATTTATTTATTTATTTAT 2100  
2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTATTTATTTATTTATTTATTTAT 2100  
2101 ATGTTGAGCA CCAATATGATATTTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 2160  
2101 ATGTTGAGCA CCAATATGATATTTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 2160  
2161 GAGAGAA CCAATTTAAACAGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
2161 GAGAGAA CCAATTTAAACAGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
2221 AAAAATGTTGTTTAAAGCTTTAAACAGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2221 AAAAATGTTGTTTAAAGCTTTAAACAGTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2281 GTGAGAA CCAAGAGAGGCTCTTAACTGATCAATGATGATGATGATGATGATGATGATGAT 2340



```

QY 1621 TCCGGGTGGTGAAGAAACAGAACTTCGCTGAGAGAACAGATGCAACAGATGTAAGGCCCC 1680
DB 1621 TCCGGGTGGTGAAGAAACAGAACTTCGCTGAGAGAACAGATGCAACAGATGTAAGGCCCC 1680
QY 1681 AAAGCTGAAGGCTTCCTCCGCAACCTTTCCGCCCCGGGTGGTATGTCGCAAGG 1740
DB 1681 AAAGCTGAAGGCTTCCTCCGCAACCTTTCCGCCCCGGGTGGTATGTCGCAAGG 1740
QY 1741 TGGCCCTGGTGGTGAAGGAGGAGAGAGAGTGGCTTCATGATGTCGTGTGTCGCGGTG 1800
DB 1741 TGGCCCTGGTGGTGAAGGAGGAGAGAGAGTGGCTTCATGATGTCGTGTGTCGCGGTG 1800
QY 1801 AATGTTCAAGATGTCGCTGTCGAGACAGAGGTGCTTCGTCGTGTCGCGGAGATGA 1860
DB 1801 AATGTTCAAGATGTCGCTGTCGAGACAGAGGTGCTTCGTCGTGTCGCGGAGATGA 1860
QY 1861 CCGAGGTGCTTTGTCGAGAGAACAGAGTGGCTTCGCGGAGGCCCCCTGAGCTTTGAT 1920
DB 1861 CCGAGGTGCTTTGTCGAGAGAACAGAGTGGCTTCGCGGAGGCCCCCTGAGCTTTGAT 1920
QY 1921 GGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 GGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 GCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1981 GCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 GACTACCAAGATTTATTTTAAACAGAGAAATGTTTAAATTAATTAATTAATTAATTA 2100
DB 2041 GACTACCAAGATTTATTTTAAACAGAGAAATGTTTAAATTAATTAATTAATTAATTA 2100
QY 2101 ATGTGGCCCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
DB 2101 ATGTGGCCCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
QY 2161 GAAGAAACATTAATAACAGTTAATGTAAGTGGAGATTTTCTTCTCTCTT 2220
DB 2161 GAAGAAACATTAATAACAGTTAATGTAAGTGGAGATTTTCTTCTCTCTT 2220
QY 2221 AAAAATGTTGTTAAGACTTTTAAACAGTGAACCCCTGTGAGCATGCTCAGATCAT 2280
DB 2221 AAAAATGTTGTTAAGACTTTTAAACAGTGAACCCCTGTGAGCATGCTCAGATCAT 2280
QY 2281 GTGAGAGACCAAGAGGCTCTTAACGTGAACAGTGTTCATGTCGTGAGATTTTCTT 2340
DB 2281 GTGAGAGACCAAGAGGCTCTTAACGTGAACAGTGTTCATGTCGTGAGATTTTCTT 2340
QY 2341 TTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2390
DB 2341 TTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2390

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
US-09-960-706-1081

Query Match          99.9%; Score 2388.4; DB 10; Length 2390;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 1 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 61 CAGTACCTTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 61 CAGTACCTTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 TCAGAGATATGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 121 TCAGAGATATGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 GCCCATGATATGTCAGTATATCCAGAGCTCAGAGCACTGCAACCTATAGGAGAGAGAG 240
DB 181 GCCCATGATATGTCAGTATATCCAGAGCTCAGAGCACTGCAACCTATAGGAGAGAGAG 240
QY 241 TGCACCTTCTTAATGAGACAGAGCTCCAGTGTATTAATTAATTAATTAATTAATTA 300
DB 241 TGCACCTTCTTAATGAGACAGAGCTCCAGTGTATTAATTAATTAATTAATTAATTA 300
QY 301 CAGCAGAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 CAGCAGAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 CTATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 CTATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 TGAAGTATGCAACCTCAATCTAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 TGAAGTATGCAACCTCAATCTAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 ACAGAGTATCTACAGTATATCCAGAGTACCTGAGAGTACCCATGAGAGAGAGAGAG 600
DB 541 ACAGAGTATCTACAGTATATCCAGAGTACCTGAGAGTACCCATGAGAGAGAGAGAG 600
QY 601 ACCTCATCTACCTCTCTACAGAGTATTCCTTACAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 ACCTCATCTACCTCTCTACAGAGTATTCCTTACAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 CAGTTAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 CAGTTAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TGGTCAACAAAGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TGGTCAACAAAGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 CTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 CCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

```

```

RESULT 3
US-09-960-706-1081
; Sequence 1081, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E;
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1081
; LENGTH: 2390
; TYPE: DNA

```

QY	961	ATTATGATGATGAGGCAATGAGCAGAGGATGGGGGAGAGAGAGCGCGGTGGAAATGGGAG	102
Db	961	ATTGATGATGATGAGGCAATGAGCAGAGGATGGGGGAGAGAGAGCGCGGTGGAAATGGGAG	1020
QY	1021	CGCTGAGAGCGAGGGTGGCTTCAATTAAGCTTGTTGGAACCCATGATGTAAGGACAGATCT	1080
Db	1021	CGCTGAGAGCGAGGGTGGCTTCAATTAAGCTTGTTGGAACCCATGATGTAAGGACAGATCT	1080
QY	1081	TGATTTAGGCCCACTGTAGATCCAGATGMAAGCTCTGACAAACAGTGCATTTATGTACA	1140
Db	1081	TGATTTAGGCCCTCTCTGTAGATCCAGATGMAAGCTCTGACAAACAGTGCATTTATGTACA	1140
QY	1141	AGGATTAATATGACATGTGACTCTAGATGATCTGGCACACTTTTAAGCAGTGGGGGT	1200
Db	1141	AGGATTAATATGACATGTGACTCTAGATGATCTGGCACACTTTTAAGCAGTGGGGGT	1200
QY	1201	TGTTAAGATGAAACAAGAACTGGGCAACCCATGATCCACATCTACCTGACAAGAAAC	1260
Db	1201	TGTTAAGATGAAACAAGAACTGGGCAACCCATGATCCACATCTACCTGACAAGAAAC	1260
QY	1261	AGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGCCAAGCTGC	1320
Db	1261	AGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGCCAAGGCTGC	1320
QY	1321	CGTGAATAGTGTGATGGGAAAGATTTCAAGGGAGCAACTTAAGTCTCCCTGTCTCG	1380
Db	1321	CGTGAATAGTGTGATGGGAAAGATTTCAAGGGAGCAACTTAAGTCTCCCTGTCTCG	1380
QY	1381	GAAAGAGCCTCCATATGAACAGATATCGGGGTGGTCTGCAACCCGTAGGGCAGAGGAT	1440
Db	1381	GAAAGAGCCTCCATATGAACAGATATCGGGGTGGTCTGCAACCCGTAGGGCAGAGGAT	1440
QY	1441	GCCACCACTCCGTGAGGTCCAGAGGCCCAAGAGGTCTTGGGAGACCATAGGATCG	1500
Db	1441	GCCACCACTCCGTGAGGTCCAGAGGCCCAAGAGGTCTTGGGAGACCATAGGATCG	1500
QY	1501	CATGGAGAGCCGTGAGAGGATATGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCG	1560
Db	1501	CATGGAGAGCCGTGAGAGGATATGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCG	1560
QY	1561	AGGGAACCCCTCTGAGGAGGAAAGCTCCAGACCCGAGCTGGAAGCTGGACGTGCCAA	1620
Db	1561	AGGGAACCCCTCTGAGGAGGAAAGCTCCAGACCCGAGCTGGAAGCTGGACGTGCCAA	1620
QY	1621	TCCGGGTTGTGAAACCAAGAACTTCGCTGAGAACAGAGTGCAACAGTGAAGGCCCC	1680
Db	1621	TCCGGGTTGTGAAACCAAGAACTTCGCTGAGAACAGAGTGCAACAGTGAAGGCCCC	1680
QY	1681	AAAGCTTGAAGGCTTCTCCGCGCACCTTTCCGCCCCCGGGTGGTGAATCGTGGCAGAGG	1740
Db	1681	AAAGCTTGAAGGCTTCTCCGCGCACCTTTCCGCCCCCGGGTGGTGAATCGTGGCAGAGG	1740
QY	1741	TGGCCCTGATGGAATGCGGGGAGAAAGGTGGCTCATATGATCGTGGTATCCCGGTGG	1800
Db	1741	TGGCCCTGATGGAATGCGGGGAGAAAGGTGGCTCATATGATCGTGGTATCCCGGTGG	1800
QY	1801	AATGTTCAAGAGTGGCCGTGTGAGAACAGAGTGGCTTCCGTGTGGCCGAGCATATGA	1860
Db	1801	AATGTTCAAGAGTGGCCGTGTGAGAACAGAGTGGCTTCCGTGTGGCCGAGCATATGA	1860
QY	1861	CCGAGGTGCTTTGGTGGAGGAAGACGAGGTGGCCCTGGGGGGCCCTCTGACCTTTGAT	1920
Db	1861	CCGAGGTGCTTTGGTGGAGGAAGACGAGGTGGCCCTGGGGGGCCCTCTGACCTTTGAT	1920
QY	1921	GGAACAGATGGGAGGGAAGAGAGGAGACGTGAGAGACTTGAAAAATATGATTAAGGGA	1980
Db	1921	GGAACAGATGGGAGGGAAGAGAGGAGACGTGAGAGACTTGAAAAATATGATTAAGGGA	1980
QY	1981	GCACCGTACAGAGCCGACAGATTCGGCCCTTACTAATATGAGAGACCCCGCAGAGCTGCATT	2040
Db	1981	GCACCGTACAGAGCCGACAGATTCGGCCCTTACTAATATGAGAGACCCCGCAGAGCTGCATT	2040
QY	2041	GACTACAGATTTATTTTAAACCAAAAAATGTTTTAATTTAATTCATATTATA	2100

Db	Query Match	Best Local Similarity	Score	Prod. No.	Indels	Gaps
Db	2041 GACTRCCGATTTATTTTTTTTAAACGAAAAATGTTTTAAATTTATATTCATATTTATA	99.9%	2388.4	DB 10;	Length 2390;	
Qy	2101 ATGTGGCCACAAACATTATGATTTATTCCTGTCTGTACTTTAGTATTTTTCACCAATTGT	100.0%	0	1;	Indels 0;	Gaps 0
Db	2101 ATGTGGCCACAAACATTATGATTTATTCCTGTCTGTACTTTAGTATTTTTCACCAATTGT					
Qy	2161 GAAAGAAACATTAAAAACAAGTTAAATGTAAGTGTGGGAGTTTCTTCTCTCTTTT					
Db	2161 GAAAGAAACATTAAAAACAAGTTAAATGTAAGTGTGGGAGTTTCTTCTCTCTTTT					
Qy	2221 AAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGAGCATCTCAGTATCAT					
Db	2221 AAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGAGCATCTCAGTATCAT					
Qy	2281 GTGGAGAACCAAGGGGCTCTTAACTGTAAACAATGTTCAATGTTGTGATGTTTTTTTT					
Db	2281 GTGGAGAACCAAGGGGCTCTTAACTGTAAACAATGTTCAATGTTGTGATGTTTTTTTT					
Qy	2341 TTTTTTTAAATTAATAATTCCAAATGTTTAAATPAAAAAATTTTTTTTTT					
Db	2341 TTTTTTTAAATTAATAATTCCAAATGTTTAAATPAAAAAATTTTTTTTTT					
RESULT 4						
US-09-873-319-717						
; Sequence 717, Application US/09873319A						
; Publication No. US20030134324A1						
GENERAL INFORMATION:						
APPLICANT: Munger, William E.						
APPLICANT: Kulkarni, Prakash						
APPLICANT: Getzenberg, Robert H.						
APPLICANT: Waga, Iwao						
APPLICANT: Yamamoto, Jun						
TITLE OF INVENTION: Identifying drugs for and diagnosis of Benign Prostatic						
TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles						
FILE REFERENCE: 44921-5029-US						
CURRENT APPLICATION NUMBER: US/09/873.319A						
CURRENT FILING DATE: 2001-06-05						
EARLIER APPLICATION NUMBER: US 60/223.323						
EARLIER FILING DATE: 2000-08-07						
NUMBER OF SEQ ID NOS: 755						
SOFTWARE: PatentIn Ver. 2.1						
SEQ ID NO 717						
LENGTH: 2390						
TYPE: DNA						
ORGANISM: Homo sapiens						
FEATURE:						
OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899						
US-09-873-319-717						
Qy	1 AGAGGAGAGCCGAGCTGTGAGAGAAAGAGAGAAAGAGAGAAAATGGCGTCCAGCATTA					
Db	1 AGAGGAGAGCCGAGCTGTGAGAGAAAGAGAGAAAGAGAGAAAATGGCGTCCAGCATTA					
Qy	61 CAGTACTATAGCCAGCTGACAGGCGACAGAGGCTTACAGTCTTACACACCGCCACAGCCAC					
Db	61 CAGTACTATAGCCAGCTGACAGGCGACAGAGGCTTACAGTCTTACACACCGCCACAGCCAC					
Qy	121 TCAAGGATATGACAGACCAACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGACA					
Db	121 TCAAGGATATGACAGACCAACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGACA					
Qy	181 GCCCATGATGTGAGCTATATCCAGAGCTCAGACCACTGCAACCTATGAGGCGAGACCGCTTA					
Db	181 GCCCATGATGTGAGCTATATCCAGAGCTCAGACCACTGCAACCTATGAGGCGAGACCGCTTA					
Qy	241 TGCAACTTCTTATGAGACAGCTCCCACTGTTATCTTCAACCTGCCCCCAGGCATTA					
Db	241 TGCAACTTCTTATGAGACAGCTCCCACTGTTATCTTCAACCTGCCCCCAGGCATTA					

Db 241 TGCACTCTTATGAGACGCTCCACTGGTTATTAATCTCAACTGCCCCCAGGAGATA 300  
Qy 301 CAGCCAGCCTGTCCAGGGGATATGCACTGGTGTCTTATGATACCACTCTCTACAGTAC 360  
Db 301 CAGCCAGCCTGTCCAGGGGATATGCACTGGTGTCTTATGATACCACTCTCTACAGTAC 360  
Qy 361 CACCAACCCAGGCTCTCTATGAGCTCAGTCTGCACTATGAGCACTCAGCTCTTATCAGG 420  
Db 361 CACCAACCCAGGCTCTCTATGAGCTCAGTCTGCACTATGAGCACTCAGCTCTTATCAGG 420  
Qy 421 CTATGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
Db 421 CTATGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
Qy 481 TGAGACTAGTCACTCAATCTAGCAGAGGGGGTTTACACAGCCGCTTAGAATATGG 540  
Db 481 TGAGACTAGTCACTCAATCTAGCAGAGGGGGTTTACACAGCCGCTTAGAATATGG 540  
Qy 541 ACAGAGTACCTACAGTATCCCGAGTACCTGGAGGCTACCCCATGAGCAGTCACTGC 600  
Db 541 ACAGAGTACCTACAGTATCCCGAGTACCTGGAGGCTACCCCATGAGCAGTCACTGC 600  
Qy 601 ACCCTCATCTACCTCTCTACAGCTATCTCTTACACAGCCGACTAGTATATGATCAG 660  
Db 601 ACCCTCATCTACCTCTCTACAGCTATCTCTTACACAGCCGACTAGTATATGATCAG 660  
Qy 661 CAGTTACTCTCAGAGAACCTATGAGGCAACCGAGGCTATGAGACAGAGTATCTA 720  
Db 661 CAGTTACTCTCAGAGAACCTATGAGGCAACCGAGGCTATGAGACAGAGTATCTA 720  
Qy 721 TGGTCAACAAAGCAGTATGGGAGCAGCCTCCCACTAGTTAACCCCACTGATC 780  
Db 721 TGGTCAACAAAGCAGTATGGGAGCAGCCTCCCACTAGTTAACCCCACTGATC 780  
Qy 781 CTACAGCCAAAGCTCCAACTATATAGCAACAGAGCAGCAGTACCGGAGCAGCAGTTC 840  
Db 781 CTACAGCCAAAGCTCCAACTATATAGCAACAGAGCAGCAGTACCGGAGCAGCAGTTC 840  
Qy 841 ATTCCGACAGAGCAACCCCACTAGTATGGGTTTATGGGCAAGAGTCTGAGGATTTTC 900  
Db 841 ATTCCGACAGAGCAACCCCACTAGTATGGGTTTATGGGCAAGAGTCTGAGGATTTTC 900  
Qy 901 CGGACAGAGAGAACCGGAGCAGTATGGGCTCTGATTAACCGGAGCAGGAGAGAGG 960  
Db 901 CGGACAGAGAGAACCGGAGCAGTATGGGCTCTGATTAACCGGAGCAGGAGAGAGG 960  
Qy 961 ATTGATCTGAGAGCAGTATGAGCAGAGGTGGGCGGAGAGAGAGCGGCTGAGATGG 1020  
Db 961 ATTGATCTGAGAGCAGTATGAGCAGAGGTGGGCGGAGAGAGAGCGGCTGAGATGG 1020  
Qy 1021 CGCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGGACCCATGATGAAGAGCCAGATCT 1080  
Db 1021 CGCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGGACCCATGATGAAGAGCCAGATCT 1080  
Qy 1081 TGATCTAGGCCCCCTCTGATGATCCAGATGAAGACTCTGACACAGTGAATTTATGACA 1140  
Db 1081 TGATCTAGGCCCCCTCTGATGATCCAGATGAAGACTCTGACACAGTGAATTTATGACA 1140  
Qy 1141 AGGATTAATGACAGTGTGACTCTAGATCTGAGCAGCTCTTTAAGCAGTGGGGT 1200  
Db 1141 AGGATTAATGACAGTGTGACTCTAGATCTGAGCAGCTCTTTAAGCAGTGGGGT 1200  
Qy 1201 TGTTAAGATGAACAGAGAACTGGGCAACCAATGATCCATCTTCTGAGCAAGAGAAC 1260  
Db 1201 TGTTAAGATGAACAGAGAACTGGGCAACCAATGATCCATCTTCTGAGCAAGAGAAC 1260  
Qy 1261 AGGAAAGCCCAAGAGGAGTCCATGAGAGAGCCCACTGAGCAGGCTGC 1320  
Db 1261 AGGAAAGCCCAAGAGGAGTCCATGAGAGAGCCCACTGAGCAGGCTGC 1320  
Qy 1321 CGTGAATGTGTTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAGTCTCTCTGCTCG 1380  
Db 1321 CGTGAATGTGTTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAGTCTCTCTGCTCG 1380

Qy 1381 GAAAGAGCTCCAAATGAACAGTATGCGGGGTGCTGCCACCCCGTGAAGGAGAGGAT 1440  
Db 1381 GAAAGAGCTCCAAATGAACAGTATGCGGGGTGCTGCCACCCCGTGAAGGAGAGGAT 1440  
Qy 1441 GCCACCAACACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTGTGGGAGACCAATGGTGTG 1500  
Db 1441 GCCACCAACACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTGTGGGAGACCAATGGTGTG 1500  
Qy 1501 CATGGAGGCTCGTGAAGAGATGAGAGAGGCTTCTCCAAAGAGACCCCGGGGTTCGCG 1560  
Db 1501 CATGGAGGCTCGTGAAGAGATGAGAGAGGCTTCTCCAAAGAGACCCCGGGGTTCGCG 1560  
Qy 1561 AGGGAACCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGAGACTGAGAGTGTCCAA 1620  
Db 1561 AGGGAACCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGAGAGTGTGTCCAA 1620  
Qy 1621 TCCGGGTGTGAAACCAAGACTTCCCTGAGAGACAGAGTGCACAGTAAAGCCCC 1680  
Db 1621 TCCGGGTGTGAAACCAAGACTTCCCTGAGAGACAGAGTGCACAGTAAAGCCCC 1680  
Qy 1681 AAAGCTGAAGGCTTCTCTCCGACACCTTTCCGCCCCGGGTGTGATCTGTGGCAGAGG 1740  
Db 1681 AAAGCTGAAGGCTTCTCTCCGACACCTTTCCGCCCCGGGTGTGATCTGTGGCAGAGG 1740  
Qy 1741 TGGCCTGGTGGCAGTGGGAGAGAGAGAGTGGCTCATGATCTGTGTGCTCCGAGTGG 1800  
Db 1741 TGGCCTGGTGGCAGTGGGAGAGAGAGAGTGGCTCATGATCTGTGTGCTCCGAGTGG 1800  
Qy 1801 AATGTTCAAGAGTGGCTGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGCAATGGA 1860  
Db 1801 AATGTTCAAGAGTGGCTGTGTGAGACAGAGTGGCTTCCGTGTGGCCGGGCAATGGA 1860  
Qy 1861 CCGAGGTGGCTTGTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGACTTGTAT 1920  
Db 1861 CCGAGGTGGCTTGTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGACTTGTAT 1920  
Qy 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGTGGAGACTGAAATAATGATTAAGGCGA 1980  
Db 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGTGGAGACTGAAATAATGATTAAGGCGA 1980  
Qy 1981 GCAACGTCAAGAGCGCAGAGATGGGCTTATGATGACAGAGACCCCGCAGAGCTGACT 2040  
Db 1981 GCAACGTCAAGAGCGCAGAGATGGGCTTATGATGACAGAGACCCCGCAGAGCTGACT 2040  
Qy 2041 GACTACCAATTTATTTTAAACAGAAATGTTTAAATTTATTAATTCATATTTATA 2100  
Db 2041 GACTACCAATTTATTTTAAACAGAAATGTTTAAATTTATTAATTCATATTTATA 2100  
Qy 2101 ATGTGGCACAATTAATGATTAATCTTGTCTGATCTTATGATTTTCAACATTTGT 2160  
Db 2101 ATGTGGCACAATTAATGATTAATCTTGTCTGATCTTATGATTTTCAACATTTGT 2160  
Qy 2161 GAAAGAAACATTAATAACAGTAAATGATGAGAGGAGTTTATTTTCTTCTCTTCT 2220  
Db 2161 GAAAGAAACATTAATAACAGTAAATGATGAGAGGAGTTTATTTTCTTCTCTTCT 2220  
Qy 2221 AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGAGCAGTCAATGAT 2280  
Db 2221 AAAAATGTTGTTTAAAGCTTTAAACAATGGGAAACCCCTGTGAGCAGTCAATGAT 2280  
Qy 2281 GTGAGAGACCAAGAGGAGGCTCTTAATCTGTAACAATGTTCTAGTGTGATGATTTT 2340  
Db 2281 GTGAGAGACCAAGAGGAGGCTCTTAATCTGTAACAATGTTCTAGTGTGATGATTTT 2340  
Qy 2341 TTTTAAATTAATAATTCCAAATGTTTAAATTAATAAAAAAAAAAAAAAAAAAAAA 2390  
Db 2341 TTTTAAATTAATAATTCCAAATGTTTAAATTAATAAAAAAAAAAAAAAAAAAAAA 2390

RESULT 5  
US-09-822-830A-410/c  
; Sequence 410, Application US/0982830A

Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fachel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howe, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulikote, Kamalakkar  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIORITY FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 410  
LENGTH: 2273  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-410

Query Match 91.2%; Score 2179.6; DB 9; Length 2273;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 2259; Conservative 0; Mismatches 9; Indels 54; Gaps 2;

QY 42 AAATGGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCGAGGCTAGAGT 101  
DB 2268 AAGGTGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCGAGGCTAGAGT 2209  
QY 102 CTACACCGCCGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 161  
DB 2208 CTACACCGCCGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 2149  
QY 162 GCTATGGAACCTTATGAGGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 221  
DB 2148 GCTATGGAACCTTATGAGGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 2089  
QY 222 CCTATGGGAGAGCGGCTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 281  
DB 2088 CCTATGGGAGAGCGGCTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 2029  
QY 282 CAATGCGCCGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 341  
DB 2028 CAATGCGCCGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1969  
QY 342 CCAACCTGCTACAGTACCAACCAAGGCTCTTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 401  
DB 1968 CCAACCTGCTACAGTACCAACCAAGGCTCTTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1909  
QY 402 CTGAGCTGCTTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 461  
DB 1908 CTGAGCTGCTTATGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1849  
QY 462 AGGATGGAACAAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 521  
DB 1848 AGGATGGAACAAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1789  
QY 522 AGCCGAGCTGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 581  
DB 1788 AGCCGAGCTGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1729  
QY 582 CGATGAGGAGTACGAGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 641  
DB 1728 CGATGAGGAGTACGAGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1669  
QY 642 CGAATGATTAAGTACGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 701  
DB 1668 CGAATGATTAAGTACGAGGCTCAGGATTAAGTACCTATAGCAAGCTGAGGCTAGAGT 1609

QY 702 ATGAGCAGGAGTACGATTAAGTACCAAGCAGCTATGAGGCGAGGCTCCACTAGT 761  
DB 1608 ATGAGCAGGAGTACGATTAAGTACCAAGCAGCTATGAGGCGAGGCTCCACTAGT 1549  
QY 762 ACCACCCCAACTGATCTTACAGGCAAGCTCAAGTCAATATAGCAAGCAGCA 821  
DB 1548 ACCACCCCAACTGATCTTACAGGCAAGCTCAAGTCAATATAGCAAGCAGCA 1489  
QY 822 GCTACGGGAGCAGAGTTCATTCAGCAGGCAAGCAGGCTAGGCTGTTATGAG 881  
DB 1488 GCTACGGGAGCAGAGTTCATTCAGCAGGCAAGCAGGCTAGGCTGTTATGAG 1429  
QY 882 AGGATGCTGAGGATTTTCCGACCAAGGAGAAAGCAGGCTAGGCTGTTATGAG 941  
DB 1428 AGGATGCTGAGGATTTTCCGACCAAGGAGAAAGCAGGCTAGGCTGTTATGAG 1369  
QY 942 GGGGAGGAGGAGGAGGATTTGATGCTGAGGAGTACAGAGTGGGCGGAGAG 1001  
DB 1368 GGGGAGGAGGAGGAGGATTTGATGCTGAGGAGTACAGAGTGGGCGGAGAG 1309  
QY 1002 GACGCGGTGAGATGAGGAGGCTGAGAGGAGGCTTCAATTAAGCTGAGACCA 1061  
DB 1308 GACGCGGTGAGATGAGGAGGCTGAGAGGAGGCTTCAATTAAGCTGAGACCA 1252  
QY 1062 TGATGAGAGCAAGATCTTATGAGGCTTACGAGGCTTACGATTAAGTACCA 1121  
DB 1251 TGATGAGAGCAAGATCTTATGAGGCTTACGAGGCTTACGATTAAGTACCA 1192  
QY 1122 ACAGTGAATTTATGATCAAGATTAATTAAGTACGATTAAGTACGATTAAGTAC 1181  
DB 1191 ACAGTGAATTTATGATCAAGATTAATTAAGTACGATTAAGTACGATTAAGTAC 1132  
QY 1182 TCTTTAAGCAGTGTGGGTTGTTAAGTAAAGCAAGAACTGGGCAACCCATGACCA 1241  
DB 1131 TCTTTAAGCAGTGTGGGTTGTTAAGTAAAGCAAGAACTGGGCAACCCATGACCA 1072  
QY 1242 TCTAATGAGCAAGGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1301  
DB 1071 TCTAATGAGCAAGGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1012  
QY 1302 CACCACTGCGCAAGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1361  
DB 1011 CACCACTGCGCAAGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 952  
QY 1362 TTAAGTCTTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1421  
DB 951 TTAAGTCTTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 892  
QY 1422 CCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1481  
DB 891 CCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 832  
QY 1482 CTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1541  
DB 831 CTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 772  
QY 1542 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1601  
DB 771 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 712  
QY 1602 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1661  
DB 711 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 652  
QY 1662 GCAACGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1721  
DB 651 GCAACGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 643  
QY 1722 GTGGTATGCTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1781  
DB 642 GTGGTATGCTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583  
QY 1782 ATGCTGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1841



Db 582 ATGTGTGTGTCCCGTGGAAATGTTTCAAGATGTCCTGTGGAGACAGATGGCTTCC 523  
Qy 1842 GTGTGTGTCGGGGCATGAGCCAGTGTGCTTTGTGTGAGGAGACAGATGGCTTGGG 1901  
Db 522 GTGTGTGTCGGGGCATGAGCCAGTGTGCTTTGTGTGAGGAGACAGATGGCTTGGG 463  
Qy 1902 GGGCCCTGTGACCTTTGATGAAACAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAG 1961  
Db 462 GGGCCCTGTGACCTTTGATGAAACAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAG 403  
Qy 1962 GAAAAATGATTAAGGAGGAGACCGCTGAGGAGGAGAGAGAGAGAGAGAGAGAGAG 2021  
Db 402 GAAAAATGATTAAGGAGGAGACCGCTGAGGAGGAGAGAGAGAGAGAGAGAGAGAG 343  
Qy 2022 GACCCCGCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2081  
Db 342 GACCCCGCAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 283  
Qy 2082 TTATTAATTCATATTTATTAATTTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2141  
Db 282 TTATTAATTCATATTTATTAATTTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 223  
Qy 2142 AGTATTTTTCACATTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2201  
Db 222 AGTATTTTTCACATTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 163  
Qy 2202 TTTTTCCTTCCTTCTTTTAAATTTGTTTAAAGCTTTAACAATGAGAGAGAGAG 2261  
Db 162 TTTTTCCTTCCTTCTTTTAAATTTGTTTAAAGCTTTAACAATGAGAGAGAGAG 103  
Qy 2262 GAGATGCTCAGTATCTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2321  
Db 102 GAGATGCTCAGTATCTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 43  
Qy 2322 GGTGTGATGTTTGT 2363  
Db 42 GGTGTGATGTTTGT 1  
RESULT 6  
US-09-822-830A-49/C  
Sequence 49, Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genentech Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulikote, Kamalakhar  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
PRIORITY FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIORITY FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 49  
LENGTH: 2176  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-49  
Query Match 89.6%; Score 2141.8; DB 9; Length 2176;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2154; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
Qy 33 AAG 92

Db 2171 AAGTGTGAG 2112  
Qy 93 GCTACAGTGTCTTACACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 152  
Db 2111 GCTACAGTGTCTTACACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2052  
Qy 153 GGAACAAAGAGATGAG 212  
Db 2051 GGAACAAAGAGATGAG 1992  
Qy 213 CCACTGCAACCTATGAG 272  
Db 1991 CCACTGCAACCTATGAG 1932  
Qy 273 ATATCTCTCAACTGCCCCAG 332  
Db 1931 ATATCTCTCAACTGCCCCAG 1872  
Qy 333 CTATATATACCACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 392  
Db 1871 CTATATATACCACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1812  
Qy 393 CATATGCACTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 452  
Db 1811 CATATGCACTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1752  
Qy 453 CAAAGCCGAGAGATGAG 512  
Db 1751 CAAAGCCGAGAGATGAG 1693  
Qy 513 GTTACAAACAG 572  
Db 1692 GTTACAAACAG 1633  
Qy 573 GAGTACACCCAG 632  
Db 1632 GAGTACACCCAG 1573  
Qy 633 CTACACAGCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
Db 1572 CTACACAGCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1513  
Qy 693 CGAGCAGCTATGAG 752  
Db 1512 CGAGCAGCTATGAG 1453  
Qy 753 CCACTAGTTACCAACCCCAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 812  
Db 1452 CCACTAGTTACCAACCCCAAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1393  
Qy 813 AGAGCAGCAGCTACCGGAG 872  
Db 1392 AGAGCAGCAGCTACCGGAG 1333  
Qy 873 TTTATGAGGAGAGAGTGTGAGAGATTTTCCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 932  
Db 1332 TTTATGAGGAGAGAGTGTGAGAGATTTTCCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1273  
Qy 933 CTGATTAACCGGAGCAG 992  
Db 1272 CTGATTAACCGGAGCAG 1213  
Qy 993 GGGAG 1052  
Db 1212 GGGAG 1153  
Qy 1053 GTGAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112  
Db 1152 GTGAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1093  
Qy 1113 ACTCTGACCAAGTGAATTTATGATCAAGATTTAAATGACAGTGTGATCTTATGATGATC 1172

Db 1092 ACTCTGCAACAGTGCATTTATGTACAGGATTAATGACAGTGTGACTTACATGATGC 1033  
 Qy 1173 TGGCAGCTTTTAAACAGTGTGGGGTGTGTAAAGATGAAACAGAGAACTGGGCAACCA 1232  
 Db 1032 TGGCAGCTTTTAAACAGTGTGGGGTGTGTAAAGATGAAACAGAGAACTGGGCAACCA 973  
 Qy 1233 TGAATCCATCTACCTGAGCAAGAAACAGAAAGCCCAAGGGGATGCCAGTGTCT 1292  
 Db 972 TGAATCCATCTACCTGAGCAAGAAACAGAAAGCCCAAGGGGATGCCAGTGTCT 913  
 Qy 1293 ATGAAGACCAACCACTGCAAGGCTCCGTGAAATGTTGATGGGAAAGATTTTCAAG 1352  
 Db 912 ATGAAGACCAACCACTGCAAGGCTCCGTGAAATGTTGATGGGAAAGATTTTCAAG 853  
 Qy 1353 GAGCAAACTTAAAGTCTCCCTGTGTGAAAGAAAGCTCCAAATGAACATATGAGGGGTG 1412  
 Db 852 GAGCAAACTTAAAGTCTCCCTGTGTGAAAGAAAGCTCCAAATGAACATATGAGGGGTG 793  
 Qy 1413 GTCTGCCACCCCTGAGGGGAGAGGAGCATGACCAACCACTCCGTGAGAGTCCAGGAGGCC 1472  
 Db 792 GTCTGCCACCCCTGAGGGGAGAGGAGCATGACCAACCACTCCGTGAGAGTCCAGGAGGCC 733  
 Qy 1473 CAGAGAGTCTGTGGGGAGCCATGAGGTGCATGAGGAGCCGTGAGAGATGAGAGGCT 1512  
 Db 732 CAGAGAGTCTGTGGGGAGCCATGAGGTGCATGAGGAGCCGTGAGAGATGAGAGGCT 673  
 Qy 1533 TCCTCTCAAGAGAACCCCGGGGTTCCTGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGC 1592  
 Db 672 TCCTCTCAAGAGAACCCCGGGGTTCCTGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGC 613  
 Qy 1593 ACCGAGCTGAGAGATGAGAGTGTCCAAATCCGGGTGTGAAACCAAGCTTCCCTGGA 1652  
 Db 612 ACCGAGCTGAGAGATGAGAGTGTCCAAATCCGGGTGTGAAACCAAGCTTCCCTGGA 553  
 Qy 1653 GAAACAGAGTCAACCAAGTGTAAAGCCCAAGCTTCCCTCCGCAACCTTTTC 1712  
 Db 552 GAAACAGAGTCAACCAAGTGTAAAGCCCAAGCTTCCCTCCGCAACCTTTTC 493  
 Qy 1713 CGGCCCCGGGTGTGATGTGAGCAGAGGTGGCCCTGTGAGATGCGGGGAGAAAGAGTG 1772  
 Db 492 CGGCCCCGGGTGTGATGTGAGCAGAGGTGGCCCTGTGAGATGCGGGGAGAAAGAGTG 433  
 Qy 1773 GGCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1832  
 Db 432 GGCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373  
 Qy 1833 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1892  
 Db 372 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313  
 Qy 1893 GGCCTGAGGGGGCCCTGGAACCTTTGATGAAACAGATGAGAGAAAGAGAGAGAGAGTG 1952  
 Db 312 GGCCTGAGGGGGCCCTGGAACCTTTGATGAAACAGATGAGAGAAAGAGAGAGAGAGTG 253  
 Qy 1953 GAGAGCTGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2012  
 Db 252 GAGAGCTGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 193  
 Qy 2013 AGATGAG 2072  
 Db 192 AGATGAG 133  
 Qy 2073 GTTTTAAATTTATATTCATATTTATATTTATATTTATATTTATATTTATATTTATATTTAT 2132  
 Db 132 GTTTTAAATTTATATTCATATTTATATTTATATTTATATTTATATTTATATTTATATTTAT 73  
 Qy 2133 CTGTACTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTTATGAT 2189  
 Db 72 CTGTACTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTTATGATTTTATGAT 16

RESULT 7  
 US-10-094-749-1074

; Sequence 1074, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: MAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SERI, NAOHITO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOKYU  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1074  
 ; LENGTH: 1988  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-094-749-1074  
 Query Match 75.7% Score 1808.4; DB 17; Length 1988;  
 Best Local Similarity 92.2% Pred. No. 0;  
 Matches 1987; Conservative 0; Mismatches 1; Indels 168; Gaps 1;  
 Db 33 AAGAGAGAAATGAGGCTCCAGGATTAAGTACCTTATAGCAAGCTGACAGGAGAGG 92  
 Qy 1 AAGAGAGAAATGAGGCTCCAGGATTAAGTACCTTATAGCAAGCTGACAGGAGAGG 60  
 Db 93 GCTACAGTGTTCACCGCCAGCCCACTCAAGATATGACAGACCAACCAAGGATATG 152  
 Qy 61 GCTACAGTGTTCACCGCCAGCCCACTCAAGATATGACAGACCAACCAAGGATATG 120  
 Db 153 GGCACAAAGCTATGAACTTATGACAGCCCACTGATGCACTATATCCAGGCTAGA 212  
 Qy 121 GGCACAAAGCTATGAACTTATGACAGCCCACTGATGCACTATATCCAGGCTAGA 180  
 Db 213 CCACTGCAACTATGAGGAGACCGGCTATGCACTTATGACAGGCTCCACTGGTT 272  
 Qy 181 CCACTGCAACTATGAGGAGACCGGCTATGCACTTATGACAGGCTCCACTGGTT 240  
 Db 273 ATACTACTCAACTGCCCCCAGGAGATACAGGAGCTGTGACAGGGGTATGCACTGATG 332  
 Qy 241 ATACTACTCAACTGCCCCCAGGAGATACAGGAGCTGTGACAGGGGTATGCACTGATG 300  
 Db 333 CTATGATACCACTGCTATGACATGACACCAACCAAGGCTCTCTATGACAGTCACTG 392  
 Qy 301 CTATGATACCACTGCTATGACATGACACCAACCAAGGCTCTCTATGACAGTCACTG 360  
 Db 393 CATATGCACTGAGCGCTTATCCAGGCTATGAGGAGAGAGAGAGAGAGAGAGAGAGAG 452  
 Qy 361 CATATGCACTGAGCGCTTATCCAGGCTATGAGGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 Db 453 CAAGACCGAGATGAGAAACAGCCCACTGAGAGATGATCAACTCAATCTATGACAGGGG 512  
 Qy 421 CA----- 422

Oy	513	GTTAACAACACCCAGCCTAGATATATGACAGAGTAACCTACAGTTATCCCAAGTAACCTG	572
Db	423	-----	422
Oy	573	GAAGCTACCCCGAGCAGCACTGACCTGACCTCCATCCCTTACCAAGCTATTCT	632
Db	423	-----AGCTATTCT	432
Oy	633	CTACACAGCCGACCTAGTTATATGATACAGACACTTACTCTCAGACAAACACCTATGGGCAC	692
Db	433	CTACACAGCCGACCTAGTTATATGATACAGACACTTACTCTCAGACAAACACCTATGGGCAC	492
Oy	693	CGAGCAGCTATGACAGCAGAGTAGCTATGGTCAACAAAGAGCTATGGGAGGAGCTC	752
Db	493	CGAGCAGCTATGACAGCAGGATAGCTATGGTCAACAAAGAGCTATGGGAGGAGCTC	552
Oy	753	CCACTAGTTACCCACCCCAACTGGATCTCTACAGCCAGCTCCAAAGTCAATATATGCCAAC	812
Db	553	CCACTAGTTACCCACCCCAACTGGATCTCTACAGCCAGCTCCAAAGTCAATATATGCCAAC	612
Oy	813	AGAGCAGCAGCTACGGGCGACAGAGTTTATTTCCACAGAACCAACCCAGTAGCATGGGTG	872
Db	613	AGAGCAGCAGCTACGGGCGACAGAGTTTATTTCCACAGAACCAACCCAGTAGCATGGGTG	672
Oy	873	TTTATGGGCGACAGAGTCTGGAGGATTTTTCGGAACACAGAGAAACCGAGACCTGATGGCC	932
Db	673	TTTATGGGCGACAGAGTCTGAGGATTTTTCGGAACACAGAGAAACCGAGACCTGATGGCC	732
Oy	933	CTGATTAACCGGGCGACGGGGAAGAGGGGATTTGATCTGTGAGGACATGACAGAGTGGGC	992
Db	733	CTGATTAACCGGGCGACGGGGAAGAGGGGATTTGATCTGTGAGGACATGACAGAGTGGGC	792
Oy	993	GGGGAGAGGAGACGGGTGGAATGGGACGCTGGAGACGAGGTGGCTTCAATAAGCTTG	1052
Db	793	GGGGAGAGGAGACGGGTGGAATGGGACGCTGGAGACGAGGTGGCTTCAATAAGCTTG	852
Oy	1053	GTGGACCCATGGATGGAAGAACCAACACTTGATCTTAGAGGCCACTGTGATTCACATGAG	1112
Db	853	GTGGACCCATGGATGGAAGAACCAACTTGATCTTAGAGGCCACTGTGATTCACATGAG	912
Oy	1113	ACTCTGACAAACAGTGCATTTATATGACAAAGATTTAAATGACAGTGTGACTCTAGATGATC	1172
Db	913	ACTCTGACAAACAGTGCATTTATATGACAAAGATTTAAATGACAGTGTGACTCTAGATGATC	972
Oy	1173	TGGCAGACTTTTAAAGCAGTGTGGGGTTGTTAAGATGAAACAAAGAACTGGGCAACCA	1232
Db	973	TGGCAGACTTTTAAAGCAGTGTGGGGTTGTTAAGATGAAACAAAGAACTGGGCAACCA	1032
Oy	1233	TGATTCACATCTACCTGGAACAAAGAAACAGAAAGCCCAAGGGGATCCCACTGTCCT	1292
Db	933	TGATTCACATCTACCTGGAACAAAGAAACAGAAAGCCCAAGGGGATCCCACTGTCCT	1092
Oy	1293	ATGAAGAACCACCCCACTGCAAGGCTGCGTGGAAATGGTTGATGGGAAAGATTTTCAAG	1352
Db	1093	ATGAAGAACCACCCCACTGCAAGGCTGCGTGGAAATGGTTGATGGGAAAGATTTTCAAG	1152
Oy	1353	GGAGCAAACTTAAAGTCTCCCTTGCTGGAAGAAAGCTTCCAATGAACAGTATGCGGGTGG	1412
Db	1153	GGAGCAAACTTAAAGTCTCCCTTGCTGGAAGAAAGCTTCCAATGAACAGTATGCGGGTGG	1212
Oy	1413	GTTCTGCCACCCCGTGAAGGGCAGAGGCAATGCCACCACTCCGTGGAAGTCCAGAGAGCC	1472
Db	1213	GTTCTGCCACCCCGTGAAGGGCAGAGGCAATGCCACCACTCCGTGGAAGTCCAGAGAGCC	1272
Oy	1473	CAGAGAGCTCTGGGGGAGCCCAATGGGTGCATATGGGAGGCGCGTGAAGGAATGAGAGGCT	1532
Db	1273	CAGAGAGCTCTGGGGGAGCCCAATGGGTGCATATGGGAGGCGCGTGAAGGAATGAGAGGCT	1332
Oy	1533	TCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGGAACCCCTCTGGAAGAGAAACGTCCAGC	1592
Db	1333	TCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGGAACCCCTCTGGAAGAGAAACGTCCAGC	1392
Oy	1593	ACCGAGCTGGAAGCTGGCACTGTCTCCCAATCCGGGTTGTGGAACCAAGACTTCCGCTGGA	1652

Db	1393	ACCGAGCTGGAGATCGGACAGTGTCCCAATCCGGGTTGTGGAAACAGAACTTCGCCTCGA	1452
Oy	1553	GAACAGATGCAACCAAGTGAAGGCCCCAAAGGCTCTCCCGCCACCCCTTC	1712
Db	1453	GAACAGAGTCAACCAAGTGAAGGCCCCAAAGGCTCTCCCGCCACCCCTTC	1512
Oy	1713	CGCCCCGGGTGGTGAATCGTGAGAGAGTGAGCCCTGTGGCATGCGGGGAGAAAGAGTG	1772
Db	1513	CGCCCCGGGTGGTGAATCGTGAGAGAGTGAGCCCTGTGGCATGCGGGGAGAAAGAGTG	1572
Oy	1773	GCCCTCATGATCGTGGTGTCGCCGGTGGAAATGTTCAAGAGTGGCCGTGTGGACAAG	1832
Db	1573	GCCCTCATGATCGTGGTGTCGCCGGTGGAAATGTTCAAGAGTGGCCGTGTGGACAAG	1632
Oy	1833	GTGGCTTCGGTGGTGGCCGGGGATGACCCGAGGTGGCTTTGGTGAAGAAAGACGAGTG	1892
Db	1633	GTGGCTTCGGTGGTGGCCGGGGATGACCCGAGGTGGCTTTGGTGAAGAAAGACGAGTG	1692
Oy	1893	GCCTCGGGGGGCCCCCTCGACCTTTGATGGACAGATGGGAGGAAGAAGAGAGAGTG	1952
Db	1693	GCCTCGGGGGGCCCCCTCGACCTTTGATGGACAGATGGGAGGAAGAAGAGAGAGTG	1752
Oy	1953	GAGGACTCGGAAAAATGATTAAGGGGAGGACCGTCAAGAGCCGACAGATCGGCCCTACT	2012
Db	1753	GAGGACTCGGAAAAATGATTAAGGGGAGGACCGTCAAGAGCCGACAGATCGGCCCTACT	1812
Oy	2013	AGATGCAAGAGACCCCGAGAGCTGCATGCACTACAGATTATTTTAAAAACAGAAAT	2072
Db	1813	AGATGCAAGAGACCCCGAGAGCTGCATGCACTACAGATTATTTTAAAAACAGAAAT	1872
Oy	2073	GTTTTAAATTTATTAATTCATTAATTTATGATGTTGGCCACAACATTAATGATTATTCCTGT	2132
Db	1873	GTTTTAAATTTATTAATTCATTAATTTATGATGTTGGCCACAACATTAATGATTATTCCTGT	1932
Oy	2133	CTGACTCTTGTATTTTTCACATTTGTGAAGAAACATTAACCAAGTTAAATGGT	2188
Db	1933	CTGACTCTTGTATTTTTCACATTTGTGAAGAAACATTAACCAAGTTAAATGGT	1988

```

      QUERY MATCH          26.4% ; Score 631; DB 18; Length 763;
      Best Local Similarity 92.8% ; Pred.No. 9..9e-151;
      Matches 661; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

      36 GAGGAAATGGCGTCCACGGATTTCAGTACTATAGCAGAAGCTGCAGGCCAGCAGGCT 95
      |||||

```



QY 64 TACCTATAGCCAGCTGCGAGCGGAGGAGGCTTCACTGCTTACACCGCCCAAGCCCACTCA 123  
DB 61 TACCTATAGCCAGCTGCGAGCGGAGGAGGCTTCACTGCTTACACCGCCCAAGCCCACTCA 120  
QY 124 AGGATATGCAAGACCAAGGCTATATGAGGCAACAAGCTATGAACTTATGAAAGCC 183  
DB 121 AGGATATGCAAGACCAAGGCTATATGAGGCAACAAGCTATGAACTTATGAAAGCC 180  
QY 184 CACTGATGTCAGCTATACCGAGGCTGAGCACTGCAACCTATGAGGCAAGCCCTATATG 243  
DB 181 CACTGATGTCAGCTATACCGAGGCTGAGCACTGCAACCTATGAGGCAAGCCCTATATG 240  
QY 244 AACTTCTTATGAGCAAGCTTCCCACTGCTTATCTATCTCCCACTGCCCCCAAGCAATAC 303  
DB 241 AACTTCTTATGAGCAAGCTTCCCACTGCTTATCTATCTCCCACTGCCCCCAAGCAATAC 300  
QY 304 CAGGCTGTCAGAGGAGTATGAGCACTGAGTCTATGATACCACTGCTATGCTATGCTATC 363  
DB 301 CAGGCTGTCAGAGGAGTATGAGCACTGAGTCTATGATACCACTGCTATGCTATGCTATC 360  
QY 364 CACCCAGGCTCTTATGAGCTCACTGCTATGATACCACTGCTATGCTATGCTATGCTATC 423  
DB 361 CACCCAGGCTCTTATGAGCTCACTGCTATGATACCACTGCTATGCTATGCTATGCTATC 420  
QY 424 TGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 483  
DB 421 TGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
QY 484 GACTAGTCAACTCACTATGAGCAAGGAGGAGTAAACAGGAGGAGGAGGAGGAGGAGGAGG 543  
DB 481 GACTAGTCAACTCACTATGAGCAAGGAGGAGTAAACAGGAGGAGGAGGAGGAGGAGGAGG 540  
QY 544 GAGTAACTAGCACTTATCCAGGAGTCTGAGGAGGAGTCCAGTATGAGGAGGAGGAGGAGG 603  
DB 541 GAGTAACTAGCACTTATCCAGGAGTCTGAGGAGGAGTCCAGTATGAGGAGGAGGAGGAGG 600  
QY 604 TCCATCTCACTCTCTCACTGAGTCTTCTCTCACTGAGGAGGAGTATGAGTATGAGGAGG 663  
DB 601 TCCATCTCACTCTCTCACTGAGTCTTCTCTCACTGAGGAGGAGTATGAGTATGAGGAGG 660  
QY 664 TTACTCTCAGAGCAACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 723  
DB 661 TTACTCTCAGAGCAACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720  
QY 724 TCAACAAAGAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 783  
DB 721 TCAACAAAGAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
QY 784 CAGGCAAGGCTCAAGTCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 843  
DB 781 CAGGCAAGGCTCAAGTCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
QY 844 CCGAGAGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 903  
DB 841 CCGAGAGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
QY 904 ACCAGAGAGAAACCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 963  
DB 901 ACCAGAGAGAAACCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960  
QY 964 TGAATCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1023  
DB 961 TGAATCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
QY 1024 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1083  
DB 1021 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
QY 1084 TCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1143  
DB 1081 TCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140

QY 1144 ATTAATATGACAGTGTGACTCTAGATGATCTGAGAGACTCTTTAAGCAGTGTGGGTTGT 1203  
DB 1141 ATTAATATGACAGTGTGACTCTAGATGATCTGAGAGACTCTTTAAGCAGTGTGGGTTGT 1200  
QY 1204 TAAAGTGAACAAAGAACTGGGCAACCCATGATCAATCTTACCTGAGCAAGAAACAG 1263  
DB 1201 TAAAGTGAACAAAGAACTGGGCAACCCATGATCAATCTTACCTGAGCAAGAAACAG 1260  
QY 1264 AAAAGCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1323  
DB 1261 AAAAGCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
QY 1324 GGAATGTTTGAATGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGGA 1383  
DB 1321 GGAATGTTTGAATGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGGA 1380  
QY 1384 GAAAGCTCCAAATGAACAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1443  
DB 1381 GAAAGCTCCAAATGAACAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440  
QY 1444 ACCACCACTCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1503  
DB 1441 ACCACCACTCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500  
QY 1504 GGGAGGCGTGAAGAGATGAGAGAGGCTTCCCTCCAAAGAGACCCCGGAGTTCAGGAG 1563  
DB 1501 GGGAGGCGTGAAGAGATGAGAGAGGCTTCCCTCCAAAGAGACCCCGGAGTTCAGGAG 1560  
QY 1564 GAAACCCCTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1623  
DB 1561 GAAACCCCTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620  
QY 1624 GGGTGTGGAACCAAGAACTTGGCTGTGAGAAACAGGTGCAACCAAGTAAAGGCCCAAA 1683  
DB 1621 GGGTGTGGAACCAAGAACTTGGCTGTGAGAAACAGGTGCAACCAAGTAAAGGCCCAAA 1680  
QY 1684 GCCTGAAGGCTTCTCTCCGCAACCTTTCCGCCCCGGGTGTGATCTGTGAGAGAGTGG 1743  
DB 1681 GCCTGAAGGCTTCTCTCCGCAACCTTTCCGCCCCGGGTGTGATCTGTGAGAGAGTGG 1740  
QY 1744 CCTGTGAGCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1803  
DB 1741 CCTGTGAGCAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800  
QY 1804 GTTCAAGAGTGGCCGTGTGAGAGCAAGAGTGGCTTCCGTGTGAGGAGGAGGAGGAGGAGG 1863  
DB 1801 GTTCAAGAGTGGCCGTGTGAGAGCAAGAGTGGCTTCCGTGTGAGGAGGAGGAGGAGGAGG 1860  
QY 1864 AGGTGCTTTGTGTGAGAGAGAGGAGTGGCCCTGTGGGGGCCCCCTGAGACTTTTGAATGA 1923  
DB 1861 AGGTGCTTTGTGTGAGAGAGAGGAGTGGCCCTGTGGGGGCCCCCTGAGACTTTTGAATGA 1920  
QY 1924 ACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1983  
DB 1921 ACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
QY 1984 CCGTCAAGAGCGGAGAGATCGGCCCTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2043  
DB 1981 CCGTCAAGAGCGGAGAGATCGGCCCTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040  
QY 2044 TACCAATTTATTTTAAACAGAAATGTTTTAATTTAATTTCAATTTTATATATG 2103  
DB 2041 TACCAATTTATTTTAAACAGAAATGTTTTAATTTAATTTCAATTTTATATATG 2100  
QY 2104 TTGGCAACAACATTTATGATTTATCTTGTCTGATCTTATGATTTTTCACATTTGTGA 2163  
DB 2101 TTGGCAACAACATTTATGATTTATCTTGTCTGATCTTATGATTTTTCACATTTGTGA 2160  
QY 2164 GAAACATTTAAACAAATTTAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2223  
DB 2161 GAAACATTTAAACAAATTTAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220  
QY 2224 AATGTTGTTTAAAGCTTTTAAACAAATGGAACCCCTGTGAGAGAGTCAATATGATGTG 2283



Db	1381	GAAGCTTC	CAATGAA	CAGTATG	CGGGGG	GGTGTCTG	CGAAC	CCCGTGA	GGGCA	GAGGCAT	ATGCC	1440
Qy	1444	ACCA	CACTCCT	GTGAG	GGTCCA	GGAGAG	CTCCAG	AGGTCT	TGGGG	GAACCCAT	TGGTAT	1503
Db	1441	ACCA	CACTCCT	GTGAG	GGTCCA	GGAGAG	CTCCAG	AGGTCT	TGGGG	GAACCCAT	TGGTAT	1500
Qy	1504	GGGAG	GGCGGT	GAGAG	ATAG	TAGGAG	GGTTCCT	CTCA	MAAG	GAACCCCGGG	GTCCCG	1563
Db	1501	GGGAG	GGCGGT	GAGAG	ATAG	TAGGAG	GGTTCCT	CTCA	MAAG	GAACCCCGGG	GTCCCG	1560
Qy	1564	GAAC	CCCCCT	CTGAG	AGAG	AAAGTTC	ACAG	CCAG	CTGAG	AGAG	CTGAG	1623
Db	1561	GAAC	CCCCCT	CTGAG	AGAG	AAAGTTC	ACAG	CCAG	CTGAG	AGAG	CTGAG	1620
Qy	1624	GGGTT	TGTG	GAAC	CAAG	AACTT	CGCT	GTGA	AAAC	AGTGT	TAAG	1683
Db	1621	GGGTT	TGTG	GAAC	CAAG	AACTT	CGCT	GTGA	AAAC	AGTGT	TAAG	1680
Qy	1684	GGCT	GAAAG	GGCTT	CGTCCG	CGCA	CCCTT	CGCG	CCCCGG	GGTGT	ATGTG	1743
Db	1681	GGCT	GAAAG	GGCTT	CGTCCG	CGCA	CCCTT	CGCG	CCCCGG	GGTGT	ATGTG	1740
Qy	1744	CCCT	TGTG	TGCA	TGCG	GGGAG	AGAG	AGTGT	GGCT	CA	TGAG	1803
Db	1741	CCCT	TGTG	TGCA	TGCG	GGGAG	AGAG	AGTGT	GGCT	CA	TGAG	1800
Qy	1804	GTTCA	GAG	AGTGT	CCGT	GTG	AGAC	CAAG	AGTGT	CTT	CGTGT	1863
Db	1801	GTTCA	GAG	AGTGT	CCGT	GTG	AGAC	CAAG	AGTGT	CTT	CGTGT	1860
Qy	1864	AGGT	GGCTT	TGTG	TGAG	AGAG	CAAG	AGTGT	GGCT	CGGG	GGG	1923
Db	1861	AGGT	GGCTT	TGTG	TGAG	AGAG	CAAG	AGTGT	GGCT	CGGG	GGG	1920
Qy	1924	ACAG	TGTG	GAAG	AGAG	AGAG	AGAG	AGAG	AGAG	AGAG	AGAG	1983
Db	1921	ACAG	TGTG	GAAG	AGAG	AGAG	AGAG	AGAG	AGAG	AGAG	AGAG	1980
Qy	1984	CCGT	TAG	AGAG	CGCAG	AGAT	CGG	CCCT	TA	AGT	GCAG	2043
Db	1981	CCGT	TAG	AGAG	CGCAG	AGAT	CGG	CCCT	TA	AGT	GCAG	2040
Qy	2044	TACCA	GATTT	ATTTT	TAA	CCAG	AAAT	GTTTT	TAA	TTT	TAA	2103
Db	2041	TACCA	GATTT	ATTTT	TAA	CCAG	AAAT	GTTTT	TAA	TTT	TAA	2100
Qy	2104	TTGG	CCA	CAAC	ATTA	TGATTA	TCCT	TGT	CTG	TA	CTTA	2160
Db	2101	TTGG	CCA	CAAC	ATTA	TGATTA	TCCT	TGT	CTG	TA	CTTA	2157

RESULT 3

CR604639

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

2236 bp

mRNA

linear

HTC 21-JUL-2004

full-length cDNA clone CS0DJ013YN08 of T cells (Jurkat cell line)

Cot 10-normalized of Homo sapiens (human).

CR604639

CR604639.1 GI:50485446

HTC; cDNA

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

Li, W.B., Gruber C., Jessee, J. and Poljansky, D.

Full-length cDNA libraries and normalization

Unpublished

Contact : Feng liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue

2 (bases 1 to 2236)

Genoscope.

Direct Submission

JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - FRANCE (E-mail: sequef@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

FEATURES  
    Source                    Location/Qualifiers  
                              1..2236  
                              /organism="Homo sapiens"  
                              /mol\_type="mRNA"  
                              /db\_xref="taxon:9606"  
                              /clone="GSD0J013YN08"  
                              /tissue\_type="T cells (Jurkat cell line) Cot  
                              10-normalized"  
                              /plasmid="pCMVSPORT\_6"

ORIGIN

Query Match	89.8%	Score 2146	DB 3	Length 2236
Best Local Similarity	97.0%	Pred. No. 0	Mismatches 0	Indels 70
Matches 2236	Conservative 0			Gaps 2
26	GAGGAGGAGGAGGAGAAATGCGCTCCACGATTTACAGTACCTATAGCCAAGCTGACAGC	85		
1	GAGGAGGAGGAGGAGGAGAAATGCGCTCCACGATTTACAGTACCTATAGCCAAGCTGACAGC	60		
86	CAGCAGGGGCTTACAGTGCTTACACCGCCCGCCCTCAAGGATATGCACAGCCACCGAG	145		
61	CAGCAGGGGCTTACAGTGCTTACACCGCCCGCCCTCAAGGATATGCACAGCCACCGAG	120		
146	GCATATGGGGCAACAAAGTATGAAACCTATGAAACAGCCCACTGATGTGAGGATATACCGAG	205		
121	GCATATGGGGCAACAAAGTATGAAACCTATGAAACAGCCCACTGATGTGAGGATATACCGAG	180		
206	GCTCAGACCACTGCAACCTATGGGACAGCCGCTATGCAACTTCTTATGACAGGCTCCC	265		
181	GCTCAGACCACTGCAACCTATGGGACAGCCGCTATGCAACTTCTTATGACAGGCTCCC	240		
266	ACTGGTTATTACTTCCAACTGCCCCCGGCGATACAGCCAGCTGTCCAGGGGTATGGC	325		
241	ACTGGTTATTACTTCCAACTGCCCCCGGCGATACAGCCAGCTGTCCAGGGGTATGGC	300		
326	ACTGGTGTATATGATACCACTGCTTCAATCAACACCAACCCAGGCGCTTCTTATGACAGT	385		
301	ACTGGTGTATATGATACCACTGCTTCAATCAACACCAACCCAGGCGCTTCTTATGACAGT	360		
386	CAGTGTGATATGGAAGTCAAGCTGCTTATCAGGCTTATGGGACAGCCAGCAGCACT	445		
361	CAGTGTGATATGGAAGTCAAGCTGCTTATCAGGCTTATGGGACAGCCAGCAGCACT	420		
446	GCACCTACAAACCGCAGATGAAACAAAGCCCACTGAGACTATGTCAACCTCAATCTAGC	505		
421	GCACCTACAAACCGCAGATGAAACAAAGCCCACTGAGACTATGTCAACCTCAATCTAGC	480		
506	ACAGGGGGTTTCAACCAAGCCCAAGCTCTAGGATATGACAGAGTAACTACAGTTATCCGAG	565		
481	ACAGGGGGTTTCAACCAAGCCCAAGCTCTAGGATATGACAGAGTAACTACAGTTATCCGAG	540		
566	GTAACCTGGAGACTACCCCATGAGCAGTCACTGCACTGCATCTCACTTCACTCTTACAGC	625		
541	GTAACCTGGAGACTACCCCATGAGCAGTCACTGCACTGCATCTTCACTCTTACAGC	600		
626	TATTCCTTACACAGCCGACTAGTTATGATCAGAGCACTTACTCTCAGACAGAACCTAT	685		
601	TATTCCTTACACAGCCGACTAGTTATGATCAGAGCACTTACTCTCAGACAGAACCTAT	660		
686	GGGCAACCAAGCAGCTATGACAGCAGAGTACTATGTTCAACAAAGCAGCTATGGGAC	745		
661	GGGCAACCAAGCAGCTATGACAGCAGAGTACTATGTTCAACAAAGCAGCTATGGGAC	720		
746	CAGCTTCCCACTAGTATCCCAACCCAACTGATTCCTACAGCAAGCTTCAAGTCAATAT	805		



Db 721 CAGCTCCCACTAGTACCCACCCCAAACTGATCTCTACAGCCAAAGCTCCAAATCATAT 780  
 QY 806 AGCCACAGACAGACAGCTACGGGACAGAGATTCAATCCGACAGAACCAACCCAGTAGC 865  
 Db 781 AGCCACAGACAGACAGCTACGGGACAGAGATTCAATCCGACAGAACCAACCCAGTAGC 840  
 QY 866 ATGGGTGTTATATGGGAGAGAGTCTGAGAGATTTTCCGACCAAGAGAGAAACCGAGATG 925  
 Db 841 ATGGGTGTTATATGGGAGAGAGTCTGAGAGATTTTCCGACCAAGAGAGAAACCGAGATG 900  
 QY 926 AGTGGCCCTGATTAACGGGGACAGGGGAGAGAGGGGATTTGATCTGAGAGCATGAGAGA 985  
 Db 901 AGTGGCCCTGATTAACGGGGACAGGGGAGAGAGGGGATTTGATCTGAGAGCATGAGAGA 960  
 QY 986 GGTGGGCGGAGAGAGAGACGGGTGGAATGGGACGGCTGAGAGAGAGGTGGCTTCAAT 1045  
 Db 961 GGTGGGCGGAGAGAGAGACGGGTGGAATGGGACGGCTGAGAGAGAGGTGGCTTCAAT 1017  
 QY 1046 AAGCTGTGGAGACCCATGATGAGAGAGACAGATCTTATCTAAGGCCCACTGTATGATCCA 1105  
 Db 1018 AAGCTGTGGAGACCCATGATGAGAGAGACAGATCTTATCTAAGGCCCACTGTATGATCCA 1077  
 QY 1106 GATGAAGACTCTGACAAAGTGCATTTATGACAAAGATTTAATGACAGTGTACTCTA 1165  
 Db 1078 GATGAAGACTCTGACAAAGTGCATTTATGACAAAGATTTAATGACAGTGTACTCTA 1137  
 QY 1166 GATGATCTGCAAGCTTTTAAAGACAGTGTGGGTTTAAAGTAAACAAGAACTGGG 1225  
 Db 1138 GATGATCTGCAAGCTTTTAAAGACAGTGTGGGTTTAAAGTAAACAAGAACTGGG 1197  
 QY 1226 CAACCCATGATCAATCTACCTGGGACAGAGAAACAGAGAAACCCAAAGCCATGATCCA 1285  
 Db 1198 CAACCCATGATCAATCTACCTGGGACAGAGAAACAGAGAAACCCAAAGCCATGATCCA 1257  
 QY 1286 GTGTCTATGAGAGACCCCACTGCAAGAGTCCCTGGAATGTTGATGGGAAAGAT 1345  
 Db 1258 GTGTCTATGAGAGACCCCACTGCAAGAGTCCCTGGAATGTTGATGGGAAAGAT 1317  
 QY 1346 TTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGGAAGAGCTCCATGAAACAGTATG 1405  
 Db 1318 TTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGGAAGAGCTCCATGAAACAGTATG 1377  
 QY 1406 CGGGGTGTGTGCAACCCCGTGAAGGAGAGAGATGCAACAACCTCCCTGAGAGTCCA 1465  
 Db 1378 CGGGGTGTGTGCAACCCCGTGAAGGAGAGAGATGCAACAACCTCCCTGAGAGTCCA 1437  
 QY 1466 GAGAGCCCAAGAGAGTCTGAGGAGAGACCAATGGGTGCAATGGGAGGCGTGGAGAGATAGA 1525  
 Db 1438 GAGAGCCCAAGAGAGTCTGAGGAGAGACCAATGGGTGCAATGGGAGGCGTGGAGAGATAGA 1497  
 QY 1526 GAGAGCTTCCCTCAAGAGAGACCCCGGAGTTCCCGAGAGAAACCCCTTGTGAGAGAGAAAC 1585  
 Db 1498 GAGAGCTTCCCTCAAGAGAGACCCCGGAGTTCCCGAGAGAAACCCCTTGTGAGAGAGAAAC 1557  
 QY 1586 GTCCACAGACCGAGCTGAGAGACTGTCCCAATCCGGGTGTGGAACCAAGAAATTC 1645  
 Db 1558 GTCCACAGACCGAGCTGAGAGACTGTCCCAATCCGGGTGTGGAACCAAGAAATTC 1617  
 QY 1646 GCTTGAAGAACAGAGTSCAACAGTGTAAAGGCCCAAGGCTGAGAGCTTCTCCGCCCA 1705  
 Db 1618 GCTTGAAGAACAGAGTSCAACAGTGTAAAGGCCCAAGGCTTCTCTCCGCCCA 1677  
 QY 1706 CCGTTTCCGCCCCCGGGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGGAGAGA 1765  
 Db 1678 CCGTTTCCGCCCCCGGGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGGAGAGA 1737  
 QY 1766 AGAGGTGGCTCATGATCTGTGTCGTCGCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGA 1825  
 Db 1738 AGAGGTGGCTCATGATCTGTGTCGTCGCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGA 1797  
 QY 1826 GACAGAGTGGCTTCCGTGTGTGCGCGGGAGATGAGACCGAGTGGCTTGTGTGAGAGAGA 1885  
 Db 1798 GACAGAGTGGCTTCCGTGTGTGCGCGGGAGATGAGACCGAGTGGCTTGTGTGAGAGAGA 1857

QY 1886 CGAGTGGCCCTGTGGGGGCCCTCGACCTTTGATGAGAACAGATGGAGAGAGAGAGA 1945  
 Db 1858 CGAGTGGCCCTGTGGGGGCCCTCGACCTTTGATGAGAACAGATGGAGAGAGAGAGA 1879  
 QY 1946 GGAAGTGAAGAGACTGGAAAAATGAGATTAAGGAGAGACCGGTGAGAGAGAGATCGG 2005  
 Db 1880 -----AGGAGAGACCGGTGAGAGAGAGATCGG 1910  
 QY 2006 CCTACTAGATGAGAGAGACCCCGAGAGTGCATTTGACATCAAGATTTATTTTAAACC 2065  
 Db 1911 CCTACTAGATGAGAGAGACCCCGAGAGTGCATTTGACATCAAGATTTATTTTAAACC 1970  
 QY 2066 AGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2125  
 Db 1971 AGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2030  
 QY 2126 TCCGTGTCTGTACTTGTATGATTTTTCACCACTTTGTGAGAGAAACATTAAACAGTTAAAT 2185  
 Db 2031 TCCGTGTCTGTACTTGTATGATTTTTCACCACTTTGTGAGAGAAACATTAAACAGTTAAAT 2090  
 QY 2186 GGTAGTGTGGAGATTTTTCCTTCTTTTAAATGAGTGTGTTAAGACTTAAAC 2245  
 Db 2091 GGTAGTGTGGAGATTTTTCCTTCTTTTAAATGAGTGTGTTAAGACTTAAAC 2150  
 QY 2246 AATGGAAACCCCTTGTGAGAGTCTCATTTGTGAGAAACCAAGAGGCTCTTAA 2305  
 Db 2151 AATGGAAACCCCTTGTGAGAGTCTCATTTGTGAGAAACCAAGAGGCTCTTAA 2210  
 QY 2306 CTGTAAACATGTTCAATGTTGTGATG 2331  
 Db 2211 CTGTAAACATGTTCAATGTTGTGATG 2236

RESULT 4  
 CR625247 2159 bp mRNA linear HTC 21-JUL-2004  
 LOCUS  
 DEFINITION  
 full-length cDNA clone CSDB005YF11 of Placenta of Homo sapiens (human).  
 CR625247.1 GI:50506054  
 VERSION  
 KEYWORDS  
 HTC; CNSLT cDNA.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 Li W.B., Gruber, C., Jesse, J. and Polyes, D.  
 TITLE  
 Full-length cDNA libraries and normalization  
 JOURNAL  
 Unpublished  
 CONTACT  
 Contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
 REMARK  
 2 (bases 1 to 2159)  
 REFERENCE  
 Genoscope.  
 AUTHORS  
 Direct Submission  
 TITLE  
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 JOURNAL  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 COMMENT  
 Location/Qualifiers  
 FEATURES  
 source  
 1..2159  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSDB005YF11"  
 /issue\_type="Placenta"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN



Db 2100 ATGTGGCCACACATTATATTCCTTCTGTCTTACTTATTTTCCACATTGT 2159  
RESULT 5  
CR610888  
LOCUS  
DEFINITION  
CR610888 2126 bp mRNA linear HTC 21-JUL-2004  
full-length cDNA clone CS0DF037Y107 of Fetal brain of Homo sapiens  
(human).  
ACCESSION  
CR610888 GI:50491695  
VERSION  
CR610888.1  
KEYWORDS  
HTC; CNSLT cDNA.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 2126)  
Li, W. B., Gruber, C., Jesses, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Peng Liang Email: liang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
Genoscope.  
2 (bases 1 to 2126)  
Genoscope.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 101 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source  
1. 2126  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF037Y107"  
/issue\_type="Fetal brain"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 88.4%; Score 2113; DB 3; Length 2126;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2126; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
QY 56 GATTACAGTACCTATGACCAAGCTGACGGCAGACAGGGCTTACAGTGTACACGGCCAG 115  
DB 1 GATTACAGTACCTATGACCAAGCTGACGGCAGACAGGGCTTACAGTGTACACGGCCAG 60  
QY 116 CCCACTCAAGATATGACAGACCAACCCAGCATATGGGCAAAAGCTATGAACTTAT 175  
DB 61 CCCACTCAAGATATGACAGACCAACCCAGCATATGGGCAAAAGCTATGAACTTAT 120  
QY 176 GGAAGACCCCACTGATGCTATACCCAGGCTTCAAGCACTGCAACTATGGGCAAGC 235  
DB 121 GGAAGACCCCACTGATGCTATACCCAGGCTTCAAGCACTGCAACTATGGGCAAGC 180  
QY 236 GGCCTAGCAACTCTTATGACAGCCCTCCCACTGGTTATCTACTCAATGGCCCCCAG 295  
DB 181 GGCCTAGCAACTCTTATGACAGCCCTCCCACTGGTTATCTACTCAATGGCCCCCAG 240  
QY 296 GCATACAGCCAGCTGCTCAAGGGGTATGCACTGTGCTTATGATACCACTGTCTAC 355  
DB 241 GCATACAGCCAGCTGCTCAAGGGGTATGCACTGTGCTTATGATACCACTGTCTAC 300  
QY 356 GTACACCAACCAAGGCTCTTATGACGCTCACTGATATGCACTGAGCTGCTTAT 415  
DB 301 GTACACCAACCAAGGCTCTTATGACGCTCACTGATATGCACTGAGCTGCTTAT 360  
QY 416 CCAAGCTATGGGAGGAGCAAGCACTGCACTGCAAGACCGCAGATGGAACAG 475

DB 361 CCAAGCTATGGGAGGAGCAAGCAAGCACTGCACTGCAAGACCGCAGATGGAACAG 420  
QY 476 CCCACTGAGACTGATCAACTCAATCTAGCAGAGGGGTATCAACCAAGCCAGCTTAGA 535  
DB 421 CCCACTGAGACTGATCAACTCAATCTAGCAGAGGGGTATCAACCAAGCCAGCTTAGA 480  
QY 536 TATGACAGAGTAACTTCACTTATCCCAAGGTACTGGGAGCTTACCCATGACGCTC 595  
DB 481 TATGACAGAGTAACTTCACTTATCCCAAGGTACTGGGAGCTTACCCATGACGCTC 540  
QY 596 ACTGACCTTCATCTCACTCCCTCAAGAGTATTCCTTCAAGAGCCGACTAGTTATGAT 655  
DB 541 ACTGACCTTCATCTCACTCCCTCAAGAGTATTCCTTCAAGAGCCGACTAGTTATGAT 600  
QY 656 CAGAGCAGTACTCTGACAGAAACCTATGGGCAACCGAGCAGTATGACAGCAGAGT 715  
DB 601 CAGAGCAGTACTCTGACAGAAACCTATGGGCAACCGAGCAGTATGACAGCAGAGT 660  
QY 716 AGCTATGCTCAAAAGCAGCTATGGGAGAGCTCCCACTTGTATCCCACTCAACT 775  
DB 661 AGCTATGCTCAAAAGCAGCTATGGGAGAGCTCCCACTTGTATCCCACTCAACT 720  
QY 776 GGATCTTACAGCCAGCTCCAACTATATAGCCACAGAGCAGCAGCTACGGGAGCAG 835  
DB 721 GGATCTTACAGCCAGCTCCAACTATATAGCCACAGAGCAGCTACGGGAGCAG 780  
QY 836 AGTTCACTCCAGACAGACCAACCCAGTAGATGGGTATTTATGGGAGAGTCTGAGGA 895  
DB 781 AGTTCACTCCAGACAGACCAACCCAGTAGATGGGTATTTATGGGAGAGTCTGAGGA 840  
QY 896 TTTTCCGAGCAGAGAGAAACCGAGCATGATGGCCCTGATTAACCGGGCAGGGAGA 955  
DB 841 TTTTCCGAGCAGAGAGAAACCGAGCATGATGGCCCTGATTAACCGGGCAGGGAGA 900  
QY 956 GGGGAGTTTGAATGTGAGGAGCATGACAGAGGGGGAGAGAGAGCGGGTGAATG 1015  
DB 901 GGGGAGTTTGAATGTGAGGAGCATGACAGAGGGGGAGAGAGAGCGGGTGAATG 960  
QY 1016 GGCAGCGCTGAGAGAGGAGGTGCTTCAATTAAGCTGTGAGCCATGATGAAGCA 1075  
DB 961 G--GCGCTGAGAGAGGAGGTGCTTCAATTAAGCTGTGAGCCATGATGAAGCA 1017  
QY 1076 GATCTTGAATCTAGGCCCACTGATATCCAGATGAAGCTCTGACCAAGTGAATTTAT 1135  
DB 1018 GATCTTGAATCTAGGCCCACTGATATCCAGATGAAGCTCTGACCAAGTGAATTTAT 1077  
QY 1136 GTTCAAGATTTAATGACAGTGTGACTTATGATCTGAGACCTTTTAAAGCTGT 1195  
DB 1078 GTTCAAGATTTAATGACAGTGTGACTTATGATCTGAGACCTTTTAAAGCTGT 1137  
QY 1196 GGGGTGTTAAGATGAACAAGAACTGGGCAACCATGATCCACTTACTGGAACA 1255  
DB 1138 GGGGTGTTAAGATGAACAAGAACTGGGCAACCATGATCCACTTACTGGAACA 1197  
QY 1256 GAAACAGAAAGCCCAAAGGCGATGCCAGTGTCTTATGAAGCCCACTGCAAG 1315  
DB 1198 GAAACAGAAAGCCCAAAGGCGATGCCAGTGTCTTATGAAGCCCACTGCAAG 1257  
QY 1316 GCTGCGGTGGAATGTTTATGAGGAAATTTTCAAGGGAAGCAACTTAAAGTCTCCTT 1375  
DB 1258 GCTGCGGTGGAATGTTTATGAGGAAATTTTCAAGGGAAGCAACTTAAAGTCTCCTT 1317  
QY 1376 GCTCGAAGAGGCTTCAATGAACAAGTATGCGGGGTGTCTGCAACCCGCTGAGGAG 1435  
DB 1318 GCTCGAAGAGGCTTCAATGAACAAGTATGCGGGGTGTCTGCAACCCGCTGAGGAG 1377  
QY 1436 GGCATCCACCACTCCGTGAGGTCCAAGAGGCCAGAGAGTCTTGGGGAGCCCATG 1495  
DB 1378 GGCATCCACCACTCCGTGAGGTCCAAGAGGCCAGAGAGTCTTGGGGAGCCCATG 1437  
QY 1496 GGTGCGATGGGAGGCGGTGAGGAGATGAGAGAGGTTTCCCTCAAGAGAGCCCGGGGT 1555

Db 1438 GGTGGATGGAGGCGGTGGAGGATGAGAGGCTTCCCTCCAGAGACCCCGGGG 1497  
 QY 1556 TCCCGAGGGAACCCCTCTGGAGAGAAACCTCCAGACCCAGCTGGAGCTGGCTGT 1615  
 Db 1498 TCCCGAGGGAACCCCTCTGGAGAGAAACCTCCAGACCCAGCTGGAGCTGGCTGT 1557  
 QY 1616 CCCAATCCGGGTTGGGAAACGAACTTGGCTGGAGACAGAGTGAACAGTGTAAAG 1675  
 Db 1558 CCCAATCCGGGTTGGGAAACGAACTTGGCTGGAGACAGAGTGAACAGTGTAAAG 1617  
 QY 1676 GCCCAAGAGCTGAAGGCTTCTCCGCAACCTTTCGCGCCCGGGTGTGATCGTGGC 1735  
 Db 1618 GCCCAAGAGCTGAAGGCTTCTCCGCAACCTTTCGCGCCCGGGTGTGATCGTGGC 1677  
 QY 1736 AGAGTGGCCCTGGTGGCATGCGGGAGAGAGAGTGGCTCATGATCGTGGTGTCC 1795  
 Db 1678 AGAGTGGCCCTGGTGGCATGCGGGAGAGAGAGTGGCTCATGATCGTGGTGTCC 1737  
 QY 1796 GGTGAATGTTCAAGAGTGGCCGTGGTGGAGACAGAGTGGCTTCCGTTGGCCGGGGC 1855  
 Db 1738 GGTGAATGTTCAAGAGTGGCCGTGGTGGAGACAGAGTGGCTTCCGTTGGCCGGGGC 1797  
 QY 1856 ATGAGACCGAGTGGCTTGGTGGAGAGACAGAGTGGCCCTGGGGGCCCCCTGAGCT 1915  
 Db 1798 ATGAGACCGAGTGGCTTGGTGGAGAGACAGAGTGGCCCTGGGGGCCCCCTGAGCT 1857  
 QY 1916 TTGATGGAACAGATGGAGAGAGAGAGAGAGAGTGGAGACCTGGAAAAATGATTA 1975  
 Db 1858 TTGATGGAACAGATGGAGAGAGAGAGAGAGAGTGGAGACCTGGAAAAATGATTA 1917  
 QY 1976 GGGAGACCGGTGAGAGAGAGAGAGAGAGTGGAGAGAGTGGAGAGAGTGGAGAGT 2035  
 Db 1918 GGGAGACCGGTGAGAGAGAGAGAGAGAGTGGAGAGAGTGGAGAGAGTGGAGAGT 1977  
 QY 2036 GCAATTGACTACAGATTATTTTAAACAGAAATGTTTAAATTTATATTCATAT 2095  
 Db 1978 GCAATTGACTACAGATTATTTTAAACAGAAATGTTTAAATTTATATTCATAT 2037  
 QY 2096 TTATTAATGTTGGCACAACATTAATTAATTTCTGTCTGTACTTATTTTTCACCA 2155  
 Db 2038 TTATTAATGTTGGCACAACATTAATTAATTTCTGTCTGTACTTATTTTTCACCA 2097  
 QY 2156 TTTGTGAAGAAACATTAACAAAGTTAA 2184  
 Db 2098 TTTGTGAAGAAACATTAACAAAGTTAA 2126

RESULT 6  
 CR619493 2103 bp mRNA linear HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS0D10761021 of Placenta Cot 25-normalized  
 DEFINITION of Homo sapiens (human).  
 ACCESSION CR619493  
 VERSION CR619493.1 GI:50500300  
 KEYWORDS HTC; CDS; EST; cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 CONTACT: Feng Liang Email: fliang@life.technet.com URL:  
 http://fulllength.invitrogen.com/InvitrogenCorporation1600  
 Faraday Avenue  
 Redwood City, CA 94063  
 Genoscope.  
 2 (bases 1 to 2103)  
 REFERENCE  
 AUTHORS Direct Submission  
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.

FEATURES  
 source  
 1..2103  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D10761021"  
 /tissue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 88.0%; Score 2103; DB 3; Length 2103;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 56 GATTACAGTACCTATGACCACTGCGAGCGAGAGGCTAGAGCTTACACCGCCAG 115  
 Db 1 GATTACAGTACCTATGACCACTGCGAGCGAGAGGCTAGAGCTTACACCGCCAG 60  
 QY 116 CCCACTCAAGATATGACAGACCAACCGCATATGAGCAACAAAGCTATGAGACTAT 175  
 Db 61 CCCACTCAAGATATGACAGACCAACCGCATATGAGCAACAAAGCTATGAGACTAT 120  
 QY 176 GGAAGCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 235  
 Db 121 GGAAGCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
 QY 236 GCGTATGACAACTTCTATGAGAGAGCTCCCACTGTTATATCTCAACTGCCCCCAG 295  
 Db 181 GCGTATGACAACTTCTATGAGAGAGCTCCCACTGTTATATCTCAACTGCCCCCAG 240  
 QY 296 GCATACAGCCAGCTCTGCAAGGAGTATGAGAGCTGCTGATGATGATGATGATGATGAT 335  
 Db 241 GCATACAGCCAGCTCTGCAAGGAGTATGAGAGCTGCTGATGATGATGATGATGATGAT 300  
 QY 356 GTACACCAACCAAGGCTCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGATGAT 415  
 Db 301 GTACACCAACCAAGGCTCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGATGAT 360  
 QY 416 CCAGCTATGAGGAG 475  
 Db 361 CCAGCTATGAGGAG 420  
 QY 476 CCAGCTATGAGGAG 535  
 Db 421 CCAGCTATGAGGAG 480  
 QY 536 TATGAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 595  
 Db 481 TATGAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 QY 596 ACTGACCTTCACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 655  
 Db 541 ACTGACCTTCACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 600  
 QY 656 CAGAGCAGTATCTTCAAGCAACCTATGAGCAACCTATGAGCAACCTATGAGCAACCTATGAG 715  
 Db 601 CAGAGCAGTATCTTCAAGCAACCTATGAGCAACCTATGAGCAACCTATGAGCAACCTATGAG 660  
 QY 716 AGCTATGATCAACAAGCAGCTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775  
 Db 661 AGCTATGATCAACAAGCAGCTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 QY 776 GATCTTACAGCCAGCTCCAGTCAATATGAGCAACAGAGCAGAGCTACGAGCAGAG 835  
 Db 721 GATCTTACAGCCAGCTCCAGTCAATATGAGCAACAGAGCAGAGCTACGAGCAGAG 780  
 QY 836 AGTTCAATTCGACAGAGCAACCCAGTACAGTGGTGTATGAGGAGAGAGTCTGAGAG 895  
 Db 781 AGTTCAATTCGACAGAGCAACCCAGTACAGTGGTGTATGAGGAGAGAGTCTGAGAG 840



QY 273 ATACTCTCCAACTGCCCCAGGCAATACAGCCAGCTGTCCAGGGGTATGCACTGGTG 332  
 DB 181 ATACTCTCCAACTGCCCCAGGCAATACAGCCAGCTGTCCAGGGGTATGCACTGGTG 240  
 QY 333 CTTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGAGCTCACTGTG 392  
 DB 241 CTTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGAGCTCACTGTG 300  
 QY 393 CATATGGCACTGAGCTGCTTATCCAGCTTATGGGAGAGCCAGCAAGCTCACTGA 452  
 DB 301 CATATGGCACTGAGCTGCTTATCCAGCTTATGGGAGAGCCAGCAAGCTCACTGA 360  
 QY 453 CAAGACCGAGATGAGAAACAGCCCACTGAGCTAGTCAACCTCATCTTATGAGCAAGG 512  
 DB 361 CAAGACCGAGATGAGAAACAGCCCACTGAGCTAGTCAACCTCATCTTATGAGCAAGG 420  
 QY 513 GTTACCAACGAGCCAGCTTATGAGATGAGCAAGTATCACTATTCCTCAAGTACTG 572  
 DB 421 GTTACCAACGAGCCAGCTTATGAGATGAGCAAGTATCACTATTCCTCAAGTACTG 480  
 QY 573 GAGCTACCCAGTACAGCTGCTGCACTGCACTCTCTTATGAGCTTATGAGCTTCT 632  
 DB 481 GAGCTACCCAGTACAGCTGCTGCACTGCACTCTCTTATGAGCTTATGAGCTTCT 540  
 QY 633 CTACACAGCCGACTAGTATGATCAGAGCAGTACTCTGAGCAAGAAACCTATGGGCAAC 692  
 DB 541 CTACACAGCCGACTAGTATGATCAGAGCAGTACTCTGAGCAAGAAACCTATGGGCAAC 600  
 QY 693 CGAGCAGCTATGAGCAGCAGAGTATGATGATGATGATGATGATGATGATGATGATG 752  
 DB 601 CGAGCAGCTATGAGCAGCAGAGTATGATGATGATGATGATGATGATGATGATGATG 660  
 QY 753 CCACTAGTTTCCCAACCCCAAACTGAGTCTTACAGCCAGCTTCAAGTCAATATAGCAAC 812  
 DB 661 CCACTAGTTTCCCAACCCCAAACTGAGTCTTACAGCCAGCTTCAAGTCAATATAGCAAC 720  
 QY 813 AGAGCAGCAGCTACGAGGAGCAGAGTATGATGATGATGATGATGATGATGATGATG 872  
 DB 721 AGAGCAGCAGCTACGAGGAGCAGAGTATGATGATGATGATGATGATGATGATGATG 780  
 QY 873 TTTATGGGAGAGTCTGAGATTTTCCGAGCCAGAGAGAAACCGAGAGCATGATGAGC 932  
 DB 781 TTTATGGGAGAGTCTGAGATTTTCCGAGCCAGAGAGAAACCGAGAGCATGATGAGC 840  
 QY 933 CTGATTAACCGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 992  
 DB 841 CTGATTAACCGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
 QY 993 GGGAGAGAGAGCGGAGTGAATGGGAGCGCTGAGAGAGGAGGAGGAGGAGGAGGAGG 1052  
 DB 901 GGGAGAGAGAGCGGAGTGAATGGGAGCGCTGAGAGAGGAGGAGGAGGAGGAGGAGG 960  
 QY 1053 GTGAGACCATGATGAGAGAGCAGATCTTGAAGCCAGCTTATGATGATGATGATG 1112  
 DB 961 GTGAGACCATGATGAGAGAGCAGATCTTGAAGCCAGCTTATGATGATGATGATG 1020  
 QY 1113 ACTCTGACCAAGTGAATTTATGATTAAGATTAATGACAGTGTATCTTATGATGATC 1172  
 DB 1021 ACTCTGACCAAGTGAATTTATGATTAAGATTAATGACAGTGTATCTTATGATGATC 1080  
 QY 1173 TGGAGAGCTTCTTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1232  
 DB 1081 TGGAGAGCTTCTTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140  
 QY 1233 TGATCCACATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1292  
 DB 1141 TGATCCACATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 QY 1293 ATGAGAGCCCAACCACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1352  
 DB 1201 ATGAGAGCCCAACCACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
 QY 1353 GAGAGAAACTTAAAGTCTCCCTGCTGAGAGAGAGGCTCAATGAGAGATGAGGAGG 1412

DB 1261 GAGAGAAACTTAAAGTCTCCCTGCTGAGAGAGAGGCTCAATGAGAGATGAGGAGG 1320  
 QY 1413 GTCTGCAACCCCGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1472  
 DB 1321 GTCTGCAACCCCGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 QY 1473 CAGAGAGCTGAG 1532  
 DB 1381 CAGAGAGCTTCTGAG 1440  
 QY 1533 TCCCTCCAG 1592  
 DB 1441 TCCCTCCAG 1500  
 QY 1593 ACCGAGCTGAG 1652  
 DB 1501 ACCGAGCTGAG 1560  
 QY 1653 GAAAGAGTGAACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712  
 DB 1561 GAAAGAGTGAACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 QY 1713 CGCCCCGAGTGAAG 1772  
 DB 1621 CGCCCCGAGTGAAG 1680  
 QY 1773 GCTCTAG 1832  
 DB 1681 GCTCTAG 1740  
 QY 1833 GTGGCTTCCGTGAG 1892  
 DB 1741 GTGGCTTCCGTGAG 1800  
 QY 1893 GGCCTGGAG 1952  
 DB 1801 GGCCTGGAG 1860  
 QY 1953 GAG 2012  
 DB 1861 GAG 1920  
 QY 2013 AGATGAG 2072  
 DB 1921 AGATGAG 1980  
 QY 2073 GTTTTAAATTTAATTTTCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2132  
 DB 1981 GTTTTAAATTTAATTTTCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2040  
 QY 2133 CTGTACTTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2192  
 DB 2041 CTGTACTTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2083

RESULT 8  
 CR594868 2093 bp mRNA linear HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS01064YE15 of Placenta Cot 25-normalized  
 DEFINITION  
 CR594868  
 ACCESSION  
 CR594868  
 VERSION  
 CR594868.1  
 KEYWORDS  
 HTC; cDNA; GI:50475675  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1. (bases 1 to 2093)  
 AUTHORS  
 Li, W.B., Gruber, C., Jessee, J., and Polyes, D.  
 TITLE  
 Full-length cDNA libraries and normalization  
 JOURNAL  
 Unpublished  
 REMARK  
 Contact : Feng Liang Email : fliang@lifetech.com URL :



http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradey Avenue

2 (bases 1 to 2093)

## REFERENCE

Genoscope.

Direct Submission

Submitted (20-JUL-2004)

Genoscope - Centre National de Sequencage :

BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

## FEATURES

Location/Qualifiers

1..2093

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GS01064YE15"

/issue\_type="Placenta Cot 25-normalized"

/plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 87.0%; Score 2080; DB 3; Length 2093;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2093; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```
QY 56 GATTACAGTACTTATGACCAAGCTGACGCGAGGAGGCTTACAGTGTCTTACACCGCCAG 115
DB 1 GATTACAGTACTTATGACCAAGCTGACGCGAGGAGGCTTACAGTGTCTTACACCGCCAG 60
QY 116 CCCACTCAAGGATATGACAGACCAAGGAGGATATGAGCAAAAGCTTATGAACTTAT 175
DB 61 CCCACTCAAGGATATGACAGACCAAGGAGGATATGAGCAAAAGCTTATGAACTTAT 120
QY 176 GGAAGGCGCACTGATGTCAGTATATCCAGGCTCAGACCACTGCAACTTATGAGCAGACC 235
DB 121 GGAAGGCGCACTGATGTCAGTATATCCAGGCTCAGACCACTGCAACTTATGAGCAGACC 180
QY 236 GCGTATGCACTTCTTATGACAGCCTCCAGCTGTTATATCTATCCAACTGCCCCAG 295
DB 181 GCGTATGCACTTCTTATGACAGCCTCCAGCTGTTATATCTATCCAACTGCCCCAG 240
QY 296 GCATACAGCCAGCTGTCAGGAGGATATGAGCACTGCTTATGATACCACTGCTTACA 355
DB 241 GCATACAGCCAGCTGTCAGGAGGATATGAGCACTGCTTATGATACCACTGCTTACA 300
QY 356 GTACCAACCAAGGAGGCTCTTATGACAGCTCAGTCTGATATGACCTCAGCCGCTTAT 415
DB 301 GTACCAACCAAGGAGGCTCTTATGACAGCTCAGTCTGATATGACCTCAGCCGCTTAT 360
QY 416 CCAAGCTATGAGGAGCAGCAGCAGCAGCAGCTTCACTCAAGACCGCAGATGAGAAAG 475
DB 361 CCAAGCTATGAGGAGCAGCAGCAGCAGCAGCTTCACTCAAGACCGCAGATGAGAAAG 420
QY 476 CCCACTGAGCTGATCACTCACTCACTTATGACAGGAGGATTAACAACGAGCCAGCTTGA 535
DB 421 CCCACTGAGCTGATCACTCACTCACTTATGACAGGAGGATTAACAACGAGCCAGCTTGA 480
QY 536 TATGACAGAGTATCACTCACTTATCCAGGATCCTGAGGCTTACCCCATGAGCAGCCAG 595
DB 481 TATGACAGAGTATCACTCACTTATCCAGGATCCTGAGGCTTACCCCATGAGCAGCCAG 540
QY 596 ACTGACCTTCATCTTACCTCTTACAGAGTATTCCTTCAACAGCCGAGTATGATATGAT 655
DB 541 ACTGACCTTCATCTTACCTCTTACAGAGTATTCCTTCAACAGCCGAGTATGATATGAT 600
QY 656 CAGAGCAGTATCTCTGAGAGAAAGCTTATGAGGAGAAAGCAGCAGCTTATGACAGCAGAT 715
DB 601 CAGAGCAGTATCTCTGAGAGAAAGCTTATGAGGAGAAAGCAGCAGCTTATGACAGCAGAT 660
QY 716 AGCTATGATCAACAAAGCAGCTATGAGCAGAGCCTCCAGTATGATCCACCCCAACT 775
DB 716 AGCTATGATCAACAAAGCAGCTATGAGCAGAGCCTCCAGTATGATCCACCCCAACT 725
```

```
DB 661 AGCTATGATCAACAAAGCAGCTATGAGCAGAGCCTCCAGTATGATCCACCCCAACT 720
QY 776 GATCTTACAGCCAGAGCTTCAAGTCAATATATGACCAAGAGCAGCAGCTTACGAGCAG 835
DB 721 GATCTTACAGCCAGAGCTTCAAGTCAATATATGACCAAGAGCAGCAGCTTACGAGCAG 780
QY 836 AGTTTATCCGAGAGAGCAAGCCAGAGAGATGAGTGTATATGAGCAGAGCTTGAAGA 895
DB 781 AGTTTATCCGAGAGAGCAAGCCAGAGAGATGAGTGTATATGAGCAGAGCTTGAAGA 840
QY 896 TTTTCCGAGCAGAGAGAGAGCCAGAGATGAGTGTATATGAGCAGAGAGAGAGAGAG 955
DB 841 TTTTCCGAGCAGAGAGAGAGCCAGAGATGAGTGTATATGAGCAGAGAGAGAGAGAG 900
QY 956 GAGGAGTTTATGATGATGAGGAGATGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 1015
DB 901 GAGGAGTTTATGATGATGAGGAGATGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 960
QY 1016 GAGGAGTTTATGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1075
DB 961 G--GAGGAGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1017
QY 1076 GATCTTATCTAGAGCCAGCTTATGATCAGATGAAAGCTTGTACCAAGTGTAAATTTAT 1135
DB 1018 GATCTTATCTAGAGCCAGCTTATGATCAGATGAAAGCTTGTACCAAGTGTAAATTTAT 1077
QY 1136 GTACAGAGATTAATATGACAGTGTGATCTTATGATGTGTGTGTGTGTGTGTGTGT 1195
DB 1078 GTACAGAGATTAATATGACAGTGTGATCTTATGATGTGTGTGTGTGTGTGTGTGT 1137
QY 1196 GGGGTGTTAAGTATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1255
DB 1138 GGGGTGTTAAGTATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
QY 1256 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1315
DB 1198 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
QY 1316 GGTGCGTGTGAATGTTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1375
DB 1258 GGTGCGTGTGAATGTTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
QY 1376 GCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435
DB 1318 GCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
QY 1436 GGCATGCAACCAACCTCCGTGAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495
DB 1378 GGCATGCAACCAACCTCCGTGAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
QY 1496 GGTGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
DB 1438 GGTGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
QY 1556 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
DB 1498 TCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
QY 1616 CCAATATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
DB 1558 CCAATATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
QY 1676 GGCCTCAAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
DB 1618 GGCCTCAAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
QY 1736 AGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
DB 1678 AGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
QY 1796 GGTGAGATGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
DB 1738 GGTGAGATGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
```



QY 1856 ATGACCCGAGTGGCTTTGGTGGAGAAAGCAGAGTGGCCCTGGGGGGCCCCCTGGAGCT 1915  
DB 1798 ATGACCCGAGTGGCTTTGGTGGAGAAAGCAGAGTGGCCCTGGGGGGCCCCCTGGAGCT 1857  
QY 1916 TTGATGGAACAGATGGAGAGAAAGAGAGAGCTGGAGAGCTGGAGAAATATGATATA 1975  
DB 1858 TTGATGGAACAGATGGAGAGAAAGAGAGAGCTGGAGAGCTGGAGAAATATGATATA 1917  
QY 1976 GGGAGAGACCGTGAAGAGAGAGAGATGGCCCTACTAGTGAAGAGAGAGAGAGAGCT 2035  
DB 1918 GGGAGAGACCGTGAAGAGAGAGAGATGGCCCTACTAGTGAAGAGAGAGAGAGAGCT 1977  
QY 2036 GCATTGACTACGATTTATTTTAAACGAGAAAGTTTAAATTTAAATTTCAATAT 2095  
DB 1978 GCATTGACTACGATTTATTTTAAACGAGAAAGTTTAAATTTAAATTTCAATAT 2037  
QY 2096 TTATATGTTGGCCACAACTATGATTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 2151  
DB 2038 TTATATGTTGGCCACAACTATGATTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 2093

RESULT 9  
CR593724 2073 bp mRNA linear HTC 21-JUL-2004  
LOCUS full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized  
DEFINITION of Homo sapiens (human).  
ACCESSION CR593724.1 GI:50474531  
VERSION CR593724.1  
KEYWORDS HTC; cDNA; full-length cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2073)  
AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 2073)  
REFERENCE 2 (bases 1 to 2073)  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a Noci-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
FEATURES  
source Location/Qualifiers  
1..2073  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1011Y005"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 86.7%; Score 2073; DB 3; Length 2073;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2073; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 GCTACAGTGTCTTACCGCCGACCACTCAAGATATGACAGACCAAGGAGCATATG 152  
DB 1 GCTACAGTGTCTTACCGCCGACCACTCAAGATATGACAGACCAAGGAGCATATG 60  
QY 153 GGCACAAAGCTATGAACTTATGAGACAGCCCACTGATGCTATATCCAGGCTGAG 212  
DB 61 GGCACAAAGCTATGAACTTATGAGACAGCCCACTGATGCTATATCCAGGCTGAG 120

QY 213 CCACCTGCACTATGGGGGAGACCGGCTATGCACTTCTTATGAGACAGGCTCCAGCTGGTT 272  
DB 121 CCACCTGCACTATGGGGGAGACCGGCTATGCACTTCTTATGAGACAGGCTCCAGCTGGTT 180  
QY 273 ATACTACTCAACTGGCCCCCAGAGCATACAGCCAGGCTCTCCAGGGGATATGGACTGGT 332  
DB 181 ATACTACTCAACTGGCCCCCAGAGCATACAGCCAGGCTCTCCAGGGGATATGGACTGGT 240  
QY 333 CTTATGATACCACTGCTTACAGTACCAACCAAGGCTCTCTATGAGAGCTCACTCTG 392  
DB 241 CTTATGATACCACTGCTTACAGTACCAACCAAGGCTCTCTATGAGAGCTCACTCTG 300  
QY 393 CATATGGCACTAGGCTGCTTATCCAGGCTATGGGAGAGAGAGAGAGAGAGAGAGAG 452  
DB 301 CATATGGCACTAGGCTGCTTATCCAGGCTATGGGAGAGAGAGAGAGAGAGAGAGAG 360  
QY 453 CAAGACCGAGAGATGAGAAAGAGCCCACTGAGACTGAGCACTCAATCTAGAGAGAG 512  
DB 361 CAAGACCGAGAGATGAGAAAGAGCCCACTGAGACTGAGCACTCAATCTAGAGAGAG 420  
QY 513 GTTACCAACGAGCCGAGCTTATGAGTATGAGACAGTATCTACAGTTATCCCAAGTACCT 572  
DB 421 GTTACCAACGAGCCGAGCTTATGAGTATGAGACAGTATCTACAGTTATCCCAAGTACCT 480  
QY 573 GAGGCTAACCCGATGAG 632  
DB 481 GAGGCTAACCCGATGAG 540  
QY 633 CTACACAGCCGACTAGTATGATCAGAGAGAGTATCTGACAGAAACCTATGGGCAAC 692  
DB 541 CTACACAGCCGACTAGTATGATCAGAGAGAGTATCTGACAGAAACCTATGGGCAAC 600  
QY 693 CGAGCAGCTATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 752  
DB 601 CGAGCAGCTATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 753 CCAGTATGATACCCAGCCCAAGCTGATCTTACAGAGAGAGAGAGAGAGAGAGAGAG 812  
DB 661 CCAGTATGATACCCAGCCCAAGCTGATCTTACAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 813 AGAGCAGCAGCTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872  
DB 721 AGAGCAGCAGCTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 873 TTTATGGGAGAGAGTCTGAGAGATTTTCCGAGACAGAGAGAGAGAGAGAGAGAG 932  
DB 781 TTTATGGGAGAGAGTCTGAGAGATTTTCCGAGACAGAGAGAGAGAGAGAGAGAG 840  
QY 933 CTGATTAACCGGGGAG 992  
DB 841 CTGATTAACCGGGGAG 900  
QY 993 GGGAGAGAGAGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052  
DB 901 GGGAGAGAGAGACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 1053 GTGAGACCATGATGAG 1112  
DB 961 GTGAGACCATGATGAG 1020  
QY 1113 ACTCTGACACAGTGCATTTATGTAACAAGAGTTAAATGACAGTGTGACTGTAGATATC 1172  
DB 1021 ACTCTGACACAGTGCATTTATGTAACAAGAGTTAAATGACAGTGTGACTGTAGATATC 1080  
QY 1173 TGGCAGACTTCTTTAAGCAGTGTGGGTTTAAAGTGAACAAGAGAGAGAGAGAGAG 1232  
DB 1081 TGGCAGACTTCTTTAAGCAGTGTGGGTTTAAAGTGAACAAGAGAGAGAGAGAGAG 1140  
QY 1233 TGATCCACATCTACCTGAGACAGAGAAAGCCAAAGGAGATGCCAGAGTGTCT 1292  
DB 1141 TGATCCACATCTACCTGAGACAGAGAAAGCCAAAGGAGATGCCAGAGTGTCT 1200

QY 1293 ATGAAAGCAACCACTGACCAAGGCTCCGCTGAAATGTTTGAATGGAAAGATTTTCAG 1352  
 DB 1201 ATGAAGACCAACCACTGACCAAGGCTCCGCTGAAATGTTTGAATGGAAAGATTTTCAG 1260  
 QY 1353 GAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAAACAGTATGGGGGTG 1412  
 DB 1261 GAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAAACAGTATGGGGGTG 1320  
 QY 1413 GTCTGCCACCCCGTGAAGGAGCAAGGCAATGCCACCACTCCGTGGAGGCTCCAGAGAGCC 1472  
 DB 1321 GTCTGCCACCCCGTGAAGGAGCAAGGCAATGCCACCACTCCGTGGAGGCTCCAGAGAGCC 1380  
 QY 1473 CAGAGAGTCTTGGGGAGCAACCATGGGTGCAATGGAGGAGCCGTGGAGAGATGAGAGAGCT 1532  
 DB 1381 CAGAGAGTCTTGGGGAGCAACCATGGGTGCAATGGAGGAGCCGTGGAGAGATGAGAGAGCT 1440  
 QY 1533 TCCCTCCAAAGAGACCCCGGGGTTCCCGAGAGAAAGCCCTTGTGAGAGAGAAAGCTCCAGC 1592  
 DB 1441 TCCCTCCAAAGAGACCCCGGGGTTCCCGAGAGAAAGCCCTTGTGAGAGAGAAAGCTCCAGC 1500  
 QY 1593 ACCGAGCTGAGACCTGCAATGTCCTCAATCCGGGTTGTGAAACAGAACTTGGCTTGA 1652  
 DB 1501 ACCGAGCTGAGACCTGCAATGTCCTCAATCCGGGTTGTGAAACAGAACTTGGCTTGA 1560  
 QY 1653 GAACAGAGTCAACAGATGTAAAGGCCCAAGGCTTCCCTCCGCAACCTTTC 1712  
 DB 1561 GAACAGAGTCAACAGATGTAAAGGCCCAAGGCTTCCCTCCGCAACCTTTC 1620  
 QY 1713 CCGCCCGGGTGTGATCGTGGCAAGAGTGCCCTGTGTGCAATGCGGGAGAGAGAGTG 1772  
 DB 1621 CCGCCCGGGTGTGATCGTGGCAAGAGTGCCCTGTGTGCAATGCGGGAGAGAGAGTG 1680  
 QY 1773 GCTCTATGATCTGTGTGTCTCCGGTGGAAATGTTCAAGAGTGCCGTGTGTGAGACAGAG 1832  
 DB 1681 GCTCTATGATCTGTGTGTCTCCGGTGGAAATGTTCAAGAGTGCCGTGTGTGAGACAGAG 1740  
 QY 1833 GTGGCTTCCGTGTGTGGCCGGGAGCATGGAACGAGGTGGCTTGGTGGAGAGAAAGCAAGTG 1892  
 DB 1741 GTGGCTTCCGTGTGTGGCCGGGAGCATGGAACGAGGTGGCTTGGTGGAGAGAGCAAGTG 1800  
 QY 1893 GCCCTGGGGGGCCCTCTGACCTTTGATGGAACAGATGAGAGAGAGAGAGAGAGTG 1952  
 DB 1801 GCCCTGGGGGGCCCTCTGACCTTTGATGGAACAGATGAGAGAGAGAGAGAGAGTG 1860  
 QY 1953 GAGGACCTGGAAGAAATGATTAAGGGAGAGACCTGTCAAGAGCCAGAGATGCGCCCTACT 2012  
 DB 1861 GAGGACCTGGAAGAAATGATTAAGGGAGAGACCTGTCAAGAGCCAGAGATGCGCCCTACT 1920  
 QY 2013 AGATGAGAGACCCCGAGAGCTGACATGACTCAAGATTTATTTTAAACAGAAAT 2072  
 DB 1921 AGATGAGAGACCCCGAGAGCTGACATGACTCAAGATTTATTTTAAACAGAAAT 1980  
 QY 2073 GTTTTAAATTTAATTTTCAATTTTAAATGTTGAGCAACATTAATTTTCTTGT 2132  
 DB 1981 GTTTTAAATTTAATTTTCAATTTTAAATGTTGAGCAACATTAATTTTCTTGT 2040  
 QY 2133 CTGTACTTTAGTATTTTTCACATTTGTGAAGA 2165  
 DB 2041 CTGTACTTTAGTATTTTTCACATTTGTGAAGA 2073  
 RESULT 10  
 CRE02561 2070 bp mRNA linear HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS0D1068YN23 of Placenta Cot 25-normalized  
 DEFINITION  
 ACCESSION CR602561  
 VERSION CR602561.1 GI:50483368  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2070)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Peng Liang Email : filiang@life.techn.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue  
 2 (bases 1 to 2070)  
 REFERENCE  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES  
 source location/Qualifiers  
 1..2070  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1068YN23"  
 /tissue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 86.1%; Score 2057; DB 3; Length 2070;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2070; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 56 GATTACAGTACTTATAGCCAGCTGACAGCGAGCGGCTTACAGTGTCTTACACCGCCAG 115  
 DB 1 GATTACAGTACTTATAGCCAGCTGACAGCGAGCGGCTTACAGTGTCTTACACCGCCAG 60  
 QY 116 CCACCTCAAGAGATATGACAGACCAACCGAGGCTATGGGCAAAAGCTTATGAACTTAT 175  
 DB 61 CCACCTCAAGAGATATGACAGACCAACCGAGGCTATGGGCAAAAGCTTATGAACTTAT 120  
 QY 176 GGAACCCCACTGATGCTTACAGGCTTACCCAGGCTCAGACCACTGCACTTATGGGAGAGC 235  
 DB 121 GGAACCCCACTGATGCTTACAGGCTTACCCAGGCTCAGACCACTGCACTTATGGGAGAGC 180  
 QY 236 GCTTATGCACTTCTTATGAGACAGCTCCCACTGTTTATCTACTCAACTGCCCCAG 295  
 DB 181 GCTTATGCACTTCTTATGAGACAGCTCCCACTGTTTATCTACTCAACTGCCCCAG 240  
 QY 296 GCTTATGAGCAAGCTTCCAGGGGTATGAGCACTGTGCTTATGATACCACTAGCTGCTACA 355  
 DB 241 GCTTATGAGCAAGCTTCCAGGGGTATGAGCACTGTGCTTATGATACCACTAGCTGCTACA 300  
 QY 356 GTCAACCAACCAAGGCTCTTATGAGCTCAAGTCACTGATGATGAGCACTAGCTGCTTAT 415  
 DB 301 GTCAACCAACCAAGGCTCTTATGAGCTCAAGTCACTGATGATGAGCACTAGCTGCTTAT 360  
 QY 416 CCAAGCTTATGGGAGAGAGCAGAGCACTGCACTTACAGAACCGAGAGATGGAACAAG 475  
 DB 361 CCAAGCTTATGGGAGAGAGCAGAGCACTGCACTTACAGAACCGAGAGATGGAACAAG 420  
 QY 476 CCACTGAGACTTATGTAACCTCAATTTAGACAGGGGGTTTACACAGAGCCAGCTTATG 535  
 DB 421 CCACTGAGACTTATGTAACCTCAATTTAGACAGGGGGTTTACACAGAGCCAGCTTATG 480  
 QY 536 TATGAGCAGAGTACTAGAGTTATCCCGAGGTACCTGGAGGCTACCCCAATGAGAGAGTC 595  
 DB 481 TATGAGCAGAGTACTAGAGTTATCCCGAGGTACCTGGAGGCTACCCCAATGAGAGAGTC 540  
 QY 596 ACTGACCTTCATCTTACCTTCACTTACAGCTATTCCTTACACAGCGGCACTAGTTATGAT 655  
 DB 541 ACTGACCTTCATCTTACCTTCACTTACAGCTATTCCTTACACAGCGGCACTAGTTATGAT 600  
 QY 656 CAGAGAGTTACTCTCAGAGAAACACTTATGGGCAACGAGAGCTATGAGACAGCAGAGT 715











Db 1318 TCAGAGAGAGAACTTTAAATGCTCTTGTCCCAAGAAAGCTCCCAATGACAGCATGGC 1377

Qy 1408 GGGTGGTCTGCAACCCCGTGAAGGACAGAGCATGCCACCACTCCGTGAGGTCCAG 1467

Db 1378 GGGAGGATGCACTCTGTGAGGAGGAGGAGGAGTCCCACTCCGTGAGGTCTGAG 1437

Qy 1468 AGGCCCCAGAGAGTCTGGGGGAGCCATGGGTCCCATGGGAGGCGGTGAGAGAAATGAGG 1527

Db 1438 TGGCCCAAGAGAGGCTCGAGAGAGCCCATGGGTCCCATGGGAGGCGGTGAGAGAGACAGAG 1497

Qy 1528 AGGCTTCCCTCAAGAGAGACCCCGGGGTTCCTCCAGAGGAAACCTCTGAGAGAGAAACGT 1587

Db 1498 GGGCTTCCCTCAAGAGAGGCGCCGAGAGCTCCAGAGAGAAACCTCTGAGAGAGAAATGT 1557

Qy 1588 CCAGCAGCAGGCTGAGAGACTGCGAGTGTCCCAATCCGGGTTTGAGAAACCAAACTTCCG 1647

Db 1558 CCAGCAGCAGGCTGAGAGACTGCGAGTGTCCCAATCCGGGCTGTGAGAAACCAAACTTCCG 1617

Qy 1648 CTGAGAGAGAGTGCAGACAGAGTAAAGGCGCCAAAGGCTTCTCCGCGCAGC 1707

Db 1618 TTGAGAGAGAGAAATGACAGAGTAAAGGCGCCCTAAGCCCGAGGAGCTTCTCCGCGCAGC 1677

Qy 1708 CTTTCCGCCCCCGGAGTGTGATGCTGAGCAGAGGTGCGCTGTGAGCATGCGGGAGAGAG 1767

Db 1678 CTTTCCGCCCCCGGAGTGTGATGCTGAGCAGAGGTGCGCTGTGAGCATGCGGGAGAGAG 1737

Qy 1768 AGGTGGCTCATGATGATGCTGAGTGTCCGGTGAAGTGTTCAGAGTGTGCGGTGTGAGA 1827

Db 1738 AGAGAGCATGAGACCGTGTGTGTCTGTGAGAGAAATGTTCAAGTGTGCGAGTGTGAGA 1797

Qy 1828 CAGAGAGTGTCTCGTGTGTGTGCGGGGACAGAGCGAGTGTGCTGTGTGTGAGAGAGAG 1887

Db 1798 CAGAGAGTGTCTCGTGTGTGTGCGGGGACAGAGCGAGTGTGCTGTGTGTGAGAGAGAG 1857

Qy 1888 AGGTGGCTCATGATGATGCTGAGTGTCCGGTGAAGTGTTCAGAGTGTGCGGTGTGAGA 1947

Db 1858 AGGTGGCTCATGATGATGCTGAGTGTCCGGTGAAGTGTTCAGAGTGTGCGGTGTGAGA 1917

Qy 1948 ACCTGAGAGACCTGAGAAATGATTAAGGCGAGCAGCGTGAAGAGCGCAGAGTGTGCGC 2007

Db 1918 ACCTGAGAGACCTGAGAAATGATTAAGGCGAGCAGCGTGAAGAGCGCAGAGTGTGCGC 1977

Qy 2008 CTACTGATGACAGAGACCCCGGAGAGTGTGATGACATGACAGATTTATTTTAAACAG 2067

Db 1978 CTACTGATGACAGAGACCCCGGAGAGTGTGATGACATGACAGATTTATTTTAAACAG 2030

Qy 2068 AAAATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2127

Db 2031 AAAATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2090

Qy 2128 CTGTGCTGACTTAG 2143

Db 2091 CTGTGCTGACTTAG 2106

RESULT 13  
AK049743 2269 bp mRNA linear HTC 03-APR-2004  
LOCUS AK049743  
DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone:CS30046A18 product:Brain sarcoma homolog, full insert sequence.  
ACCESSION AK049743 GI:26093614  
VERSION AK049743.1  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253  
PUBMED 10349636  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20493374  
PUBMED 11042159  
REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, S., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE  
AUTHORS 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE  
AUTHORS 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Iehi, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ono, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeuchi, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
URL: http://genome.gsc.riken.jp/ Location: Qualifiers  
FEATURES  
source  
1. 2269  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:CS30046A18"  
/db\_xref="taxon:10090"  
/clone="CS30046A18"  
/issue\_type="spinal cord"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"











GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:41:37 ; Search time 8501.87 Seconds  
(without alignments)  
11541.209 Million cell updates/sec

Title: US-10-791-017A-1\_COPY\_8\_2032

Perfect score: 2025  
Sequence: 1 gacggcagcttgagagaaga.....agatgcagagaccgccgaga.2025

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Genbank1: \*  
1: gb\_ba: \*  
2: gb\_mt: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_at: \*  
12: gb\_gy: \*  
13: gb\_un: \*  
14: gb\_vl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2025	100.0	2390	6	CQ0867364
2	2023.4	99.9	2390	6	AX411125
3	2023.4	99.9	2390	6	HSEWS
4	2011.4	99.3	2371	6	A36460
5	2011.4	99.3	2371	6	AR080100
6	2008	99.2	2182	9	BC004817
7	2006	99.1	2189	9	CR456490
8	1979	97.7	2164	9	BC072442
9	1975.4	97.6	2364	9	BC011048
10	1969.4	97.3	1971	12	BT007796
11	1966	97.1	2189	9	AK056309
12	1964.4	97.0	2326	6	C0730352
13	1884.4	93.1	193559	2	AC137500
14	1879.6	92.8	155815	2	AC018774
15	1879.6	92.8	180718	9	AL596087
16	1879.6	92.8	182501	2	AC011221
17	1798	88.8	2189	10	BC068226
18	1787.2	88.3	2188	6	AX305537
19	1787.2	88.3	2188	10	MMEMS

20	1785.6	88.2	2396	5	BC083960
21	1731.6	85.5	132906	2	AC109802
22	1731.6	85.5	163104	2	AC110672
23	1652.4	81.6	1988	6	AX714390
24	1652.4	81.6	1988	9	AK056681
25	1625.8	80.3	2026	6	C0850483
26	1625.8	80.3	2026	9	AK127624
27	1517.6	74.9	226400	2	AC128482
28	1517.6	74.9	235241	2	AC098231
29	1517.6	74.9	263925	2	AC106522
30	1428	70.5	1783	6	C0721057
31	1383.4	68.3	221444	2	AC113313
32	1373.8	67.8	247757	2	AC121282
33	1302.4	64.3	2440	5	AJ719366
34	1223	60.4	221924	2	AC105889
35	1218	60.1	275105	2	AC118907
36	1168.8	57.7	270337	2	AC099212
37	1014.4	50.1	7293	9	BSM08920
38	1010.8	49.9	1303	9	BC000527
39	979	48.3	1864	9	AF254086
40	979	48.3	2179	9	AF254087
41	979	48.3	2252	9	AF254088
42	976.4	48.2	2473	9	AK026270
43	829.4	41.0	2107	9	HSU35622
44	819	40.4	2412	6	I66247
45	801.2	39.6	2665	5	BC063928

#### ALIGNMENTS

RESULT 1  
LOCUS CQ0867364 2390 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 1 from Patent EP1455190.  
ACCESSION CQ0867364  
VERSION CQ0867364.1 GI:51997589  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 Odendorf, M. and Wolf, S.  
AUTHORS Modulation of the activity of nuclear receptors via EMS  
TITLE Patent: EP 1455190-A 1 08-SEP-2004;  
JOURNAL Schering Aktiengesellschaft (DB)

#### FEATURES

source

CDS

1..2390  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
44..2014  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAH33891.1"  
/db\_xref="GI:51997590"  
/translation="MASTDYSTYSQAHAQGYSAVYAPOTGQAOTTAAGQSYGY  
GQPTDVSTQAOTTAAGYVQAAVATSGPPTGATTPTAPOAVGAPVQVGTGAVDTT  
ATTTTQASVTAQASAVGTQAVGQQAAPTPRPPDGNKPTESISOPSSSYGNO  
PSLGTQASVSTPQVPGSTPMPQVTPAPSPPTPSYSTSTPTSTIDSSSQNTYGGPS  
STGQSSSYGQSSYGOPTSPSTPPTQSYSAVSQSSSYGQSSSYFRDHPSSNG  
VYQESGSGFSGPENSMSGPDNRGRGDFGMSRGGGRGAGMSAGRGGFN  
KPGPMDEGPDLDLGPVDPEDSDNSALVYQGLNDSVTLIDLADFPKCGVYKMR  
TQGPMTIYLDEKTRKGDATVSYEDPPTAKAAVWFPDGLDQSGKLKPLARKKP  
NMSMRGLPRLRGRGMPPLRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG  
PSSGAVGVRAGDMQCPNCGCAQNTAMTECHQCAVPRBGLTPPPPPPGDRRG  
GPGMRGGRGLMDRGPGMGFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG  
LMEQWGRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG

#### ORIGIN

Query Match 100.0%; Score 2025; DB 6; Length 2390;  
Best Local Similarity 100.0%; Pred. No. 0;



SOURCE	ORGANISM
1	Homo sapiens (human)
2	Homo sapiens
3	Euteleostomi
4	Euteleostomi
5	Euteleostomi
6	Euteleostomi
7	Euteleostomi
8	Euteleostomi
9	Euteleostomi
10	Euteleostomi
11	Euteleostomi
12	Euteleostomi
13	Euteleostomi
14	Euteleostomi
15	Euteleostomi
16	Euteleostomi
17	Euteleostomi
18	Euteleostomi
19	Euteleostomi
20	Euteleostomi
21	Euteleostomi
22	Euteleostomi
23	Euteleostomi
24	Euteleostomi
25	Euteleostomi
26	Euteleostomi
27	Euteleostomi
28	Euteleostomi
29	Euteleostomi
30	Euteleostomi
31	Euteleostomi
32	Euteleostomi
33	Euteleostomi
34	Euteleostomi
35	Euteleostomi
36	Euteleostomi
37	Euteleostomi
38	Euteleostomi
39	Euteleostomi
40	Euteleostomi
41	Euteleostomi
42	Euteleostomi
43	Euteleostomi
44	Euteleostomi
45	Euteleostomi
46	Euteleostomi
47	Euteleostomi
48	Euteleostomi
49	Euteleostomi
50	Euteleostomi
51	Euteleostomi
52	Euteleostomi
53	Euteleostomi
54	Euteleostomi
55	Euteleostomi
56	Euteleostomi
57	Euteleostomi
58	Euteleostomi
59	Euteleostomi
60	Euteleostomi
61	Euteleostomi
62	Euteleostomi
63	Euteleostomi
64	Euteleostomi
65	Euteleostomi
66	Euteleostomi
67	Euteleostomi
68	Euteleostomi
69	Euteleostomi
70	Euteleostomi
71	Euteleostomi
72	Euteleostomi
73	Euteleostomi
74	Euteleostomi
75	Euteleostomi
76	Euteleostomi
77	Euteleostomi
78	Euteleostomi
79	Euteleostomi
80	Euteleostomi
81	Euteleostomi
82	Euteleostomi
83	Euteleostomi
84	Euteleostomi
85	Euteleostomi
86	Euteleostomi
87	Euteleostomi
88	Euteleostomi
89	Euteleostomi
90	Euteleostomi
91	Euteleostomi
92	Euteleostomi
93	Euteleostomi
94	Euteleostomi
95	Euteleostomi
96	Euteleostomi
97	Euteleostomi
98	Euteleostomi
99	Euteleostomi
100	Euteleostomi

FEATURES	Location/Qualifiers
source	1. .2390

```

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. X66899"

```

**ORIGIN**

Query Match	99.9%	Score 2023.4	DB 6	Length 2390
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2024; Conservative	0	Mismatches 1	Indels 0	Gaps 0

OY	1	TACGGACCGTTGAGAGAACGAGAGGAAAGGAGAAATAATGGCGTCCACGGATTTCAGATACC	60
Db	8	GACGGACGTTGAGAGAACGAGAGGAAAGGAGAAATAATGGCGTCCACGGATTTCAGATACC	67
OY	61	TATAGCCAAAGCTGCAGCGCAGCAGAGGGCTACAGTGTCTTACACCGCCAGCCCACTCAAGGA	120
Db	68	TATAGCCAAAGCTGCAGCGCAGCAGAGGGCTACAGTGTCTTACACCGCCAGCCCACTCAAGGA	127
OY	121	TATGCAACAGACCAACCCAGGCATATGCGGCAACAAAGCTATGGAACTTATGGAACGCCCACT	180
Db	128	TATGCAACAGACCAACCCAGGCATATGCGGCAACAAAGCTATGGAACTTATGGAACGCCCACT	187
OY	181	GATGTCAGCTATATCCCAAGGCTCAGAACCACTGCAACCTATGCGGAGACCGGCTTATGCAACT	240
Db	188	GATGTCAGCTATATCCCAAGGCTCAGAACCACTGCAACCTATGCGGAGACCGGCTTATGCAACT	247
OY	241	TCTTATGGAACAGCCCTCCCACTGGTTATCTACTCCAACTGCCGCCAGGCATACAGCCAG	300
Db	248	TCTTATGGAACAGCCCTCCCACTGGTTATCTACTCCAACTGCCGCCAGGCATACAGCCAG	307
OY	301	CCTGTCAGAGGGGTATGGAACCTGGTGTCTTATGATACCAACACTGCTACAGTCACACCAACC	360
Db	308	CCTGTCAGAGGGGTATGGAACCTGGTGTCTTATGATACCAACAACCTGCTACAGTCACACCAACC	367
OY	361	CAGGCGCTCCTATGAGAGCTCAAGTCTGATATGGAACCTCAGCGTGTCTTATCCAGGCTTATGGG	420
Db	368	CAGGCGCTCCTATGAGAGCTCAAGTCTGATATGGAACCTCAGCGTGTCTTATCCAGGCTTATGGG	427
OY	421	CAGCAGCCAGCAGCACTGCACCTTACAAAGACCGCAGATGGAACAAAGCCCACTGAGACT	480
Db	428	CAGCAGCCAGCAGCACTGCACCTTACAAAGACCGCAGATGGAACAAAGCCCACTGAGACT	487
OY	481	AGTCAACCTCAATCTTAGCAACAGGGGGTTTCAACACAGCCCAAGCCTTAGATATGGAACAGAGT	540
Db	488	AGTCAACCTCAATCTTAGCAACAGGGGGTTTCAACACAGCCCAAGCCTTAGATATGGAACAGAGT	547
OY	541	AACTACAGATTATCCCCAGGTACTCTGGAGACTACCCATGACGCACTGCACCTCCCA	600
Db	548	AACTACAGATTATCCCCAGGTACTCTGGAGACTACCCATGACGCACTGCACCTCCCA	607
OY	601	TCTTACCGCTCTTACAGCACTATTCTCTTACACAGCCGACTAGTTATGATCAGACAGATTAC	660
Db	608	TCTTACCGCTCTTACAGCACTATTCTCTTACACAGCCGACTAGTTATGATCAGACAGATTAC	667
OY	661	TCTCAGCAGAACCACTATGCGGCAACCGACAGCTTATGACAGCAGAGTAGTATGTTCAA	720
Db	668	TCTCAGCAGAACCACTATGCGGCAACCGACAGCTTATGACAGCAGAGTAGTATGTTCAA	727
OY	721	CAAAACAGCTATGCGGAGAGAGCTCCCACTAGTTACCCACCCCAACTGATATCTTACAGC	780
Db	728	CAAAACAGCTATGCGGAGAGAGCTCCCACTAGTTACCCACCCCAACTGATATCTTACAGC	787

QY	761	CAAGTCCAACTCAATATATAGCCAAACAGACGACAGCTACGGGCGACAGATTCCGAC	840
Db	768	CAAGCTCCAACTCAATATATAGCCAAACAGACGACAGCTACGGGCGACAGATTCCGAC	847
QY	841	CAGAGCAACCCCAAGACATGGGGTGTTTATGGGAGAGGTGTGAGAGATTTTCCGGACCA	900
Db	848	CAGAGCAACCCCAAGACATGGGGTGTTTATGGGAGAGGTGTGAGAGATTTTCCGGACCA	907
QY	901	GGAGAGAACCCGAGACATAGATGAGGCTTGATTAACCGGGGACAGGGGAGAGGGGGATTTGAT	960
Db	908	GGAGAGAACCCGAGACATAGATGAGGCTTGATTAACCGGGGACAGGGGAGAGGGGGATTTGAT	967
QY	961	CGTGGAGGCATGACAGAGGTGGGGGGGAGAGAGACGGGTGAAATGGGACGCGCTGGA	1020
Db	968	CGTGGAGGCATGACAGAGGTGGGGGGGAGAGAGACGGGTGAAATGGGACGCGCTGGA	1027
QY	1021	GAGCGAGGTGGCTTCAATTAAGCCGTGGTGGACCCATGGAATGAAAGAACCAAGATCTTGAATCTA	1080
Db	1028	GAGCGAGGTGGCTTCAATTAAGCCGTGGTGGACCCATGGAATGAAAGAACCAAGATCTTGAATCTA	1087
QY	1081	GGCCCACTGTAGATTCAGATGAAAGACTCTGACAAAGTGCAATTTATGTCAAGATTA	1140
Db	1088	GGCCCTCTGTAGATTCAGATGAAAGACTCTGACAAAGTGCAATTTATGTCAAGATTA	1147
QY	1141	AATGACAGTGTGACTATAGATGATCTGGCAGACTTTTAAAGCAGTGTGGGTTGTTAAG	1200
Db	1148	AATGACAGTGTGACTATAGATGATCTGGCAGACTTTTAAAGCAGTGTGGGTTGTTAAG	1207
QY	1201	ATGAACAAAGAAATCTGGGCAACCCATGATCCACATTAACCTTGAACAAGAAACAGAAAG	1260
Db	1208	ATGAACAAAGAAATCTGGGCAACCCATGATCCACATTAACCTTGAACAAGAAACAGAAAG	1267
QY	1261	CCCAAAGGCGATGCGCACAGTGTCTTATGAAGAACCCACCACTGCAAGAGCTGCGGTGAA	1320
Db	1268	CCCAAAGGCGATGCGCACAGTGTCTTATGAAGAACCCACCACTGCAAGAGCTGCGGTGAA	1327
QY	1321	TGTTTGTGAGGAAAGATTTTCAAGGAGACAAACTTAAGTCTCTCTTGTGGAAGAG	1380
Db	1328	TGTTTGTGAGGAAAGATTTTCAAGGAGACAAACTTAAGTCTCTCTTGTGGAAGAG	1387
QY	1381	CTTCCAAATGAACAGTATACGGGGGTGTGTGCAACCCCGTGAAGGCGAAGGCAATGCCACCA	1440
Db	1388	CTTCCAAATGAACAGTATACGGGGGTGTGTGCAACCCCGTGAAGGCGAAGGCAATGCCACCA	1447
QY	1441	CCAACTCGGTGAGGTCGAGAGGCTCGAGAGAGTCTTGAGGGGACCCATGGGTGCGATGGGA	1500
Db	1448	CCAACTCGGTGAGGTCGAGAGGCTCGAGAGAGTCTTGAGGGGACCCATGGGTGCGATGGGA	1507
QY	1501	GGCGGTGAGAGATAGAGAGAGCTTCCCTTCMAAGAGACCCCGGGGTTCCTCGAGGAAAC	1560
Db	1508	GGCGGTGAGAGATAGAGAGAGCTTCCCTTCMAAGAGACCCCGGGGTTCCTCGAGGAAAC	1567
QY	1561	CCCTCTGAGAGAGAAACGTCCAGACCCGAGCTGGAACCTGGCAGTGTCCCAATCCGGGT	1620
Db	1568	CCCTCTGAGAGAGAAACGTCCAGACCCGAGCTGGAACCTGGCAGTGTCCCAATCCGGGT	1627
QY	1621	TGTGGAAACCGAAACTTGCCTGTGAGAAACAGAGTGCAACAGTGTAAAGGCCCCCAAGCCT	1680
Db	1628	TGTGGAAACCGAAACTTGCCTGTGAGAAACAGAGTGCAACAGTGTAAAGGCCCCCAAGCCT	1687
QY	1681	GAAAGCTTCTCCCGCACCCCTTTCCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCCCT	1740
Db	1688	GAAAGCTTCTCCCGCACCCCTTTCCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCCCT	1747
QY	1741	GGTGGCATGCGGGAGAGAAAGGTGGCTCTCATGATGTGTGTGTCCCGGTGCAATGTTTC	1800
Db	1748	GGTGGCATGCGGGAGAGAAAGGTGGCTCTCATGATGTGTGTGTCCCGGTGCAATGTTTC	1807
QY	1801	AGAGGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGGCCGGGGACATGAAACGAGGT	1860
Db	1808	AGAGGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGGCCGGGGACATGAAACGAGGT	1867
QY	1861	GGCTTTGTGTGAGAGAGACGAGGTGGGCTTGAGGGGGGCCCTCTGACCTTTGATGGAACAG	1920



Db	1868	GGCTTTGTGGAGGAAGCGAGGTCGGCCCTGGGGGGCCCCCTGGACCTTTGATGGAAcAG	1927
Qy	1921	ATTGGAGAGAAAGAAAGAGAGCGTGGAGGACCTGGAAAAAAATGGATTAAGGAGACACCGT	1980
Db	1928	ATGGAGAGAAAGAAAGAGAGACGTGGAGACCTGGAAAAAAATGGATTAAGGAGACACCGT	1987
Qy	1981	CAGAGAGCGCAAGATCGGCCTTAATAATGGAGAGACCCGGCAAA	2025
Db	1988	CAGAGAGCGCAAGATCGGCCTTAATAATGGAGAGACCCGGCAAA	2032

LOCUS	SIZE	STRUCTURE	REFERENCE
HSEWS	2390 bp	linear	PRI 28-JUN-1995

DEFINITION	H. sapiens EWS mRNA.
ACCESSION	X66899
VERSION	X66899.1 GI:547565
KEYWORDS	RNA binding protein.
SOURCE	Homo sapiens (human).
ORGANISM	Homo sapiens

**REFERENCE**

**AUTHORS**

**1**

Delaire, O., Zucman, J., Plougastel, B., Desmazes, C., Melot, T.,  
Tran, A., Tubakcioglu, S., Fournier, M., Laroche, P., Laroche, A.  
et al. (2019) "The Role of the State in the Development of the  
French Economy." *Journal of Economic Surveys*, vol. 70, pp. 1-48.

**TITLE** Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours  
**JOURNAL** *Nature* 359 (6391), 162-165 (1992)  
**MEDLINE** 92396239

TITLE Direct Submission  
JOURNAL Submitted (26-MAY-1992) O. Delattre, Lab. de Genet. des Tumeurs.  
Inst. Curie, 26 rue D'Ulm, 75231 Paris Cedex, FRANCE  
COMMENT On Sep 23, 1994 this sequence version replaced gi:31279.

COMMENT	On Sep 23, 1994 this sequence version replaced g1:31279.
FEATURES	Location/Qualifiers
SOURCE	1. .2350

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="22q12"
/clone="BPIAC5"
/tissue_type="brain"
/clone_id="cdNA_ StrataGene 936206"
/dev_stage="foetus"
1..2390
gene
/gene="EWS"
44..2014
CDS

```

```

/gene="BWS"
/codon_start=1
/product="RNA binding protein"
/protein_id="CAA47350.1"
/db_xref="GI:31280"
/db_xref="GOA:001844"
/db_xref="UniProt/Swiss-Prot:Q01844"
/translation="MASTDIATVTSQAAAGQGSATAPATPTGQVATTAQVAGQGSXGTA
GDPDVSYTAQATTAATGATGATVATVSYGQPVQVATPTTAQKASVQVQGVGAVDTTIT
ATVTTTQASTAAQSAVGTQAPVATGQAPAAATATPRBDKATPTSPQSTGTYN
PSLGGQSNYSYPOVPSYPMQVATPAPSYPTPTSYSPQVTSYDSSYDQNTTGGP
SYGQSSYAGQSSYGQAPPTSYVPQTGYSQAAPSYGQSSSYGQSSSFPRDHPESMQ
VYGGSGSGPFGSGENRSMGSDNMRGSGFGPGRGMSGSGGSGAGSGERGSF
KRGSGMDSGPDLIDGPPVDPDDESDNSAIYVQGLNLSYTLIDLDAPFPGCGVKKNRK
TGQPRNTHYLDKETKPKGDAATVSYEDPPTAKANEMWDGKDFQSGTKVSLARKPP
MWSKNGILPFRBGRMPPPLKRGPGGPGGPGMGKGRGDRGFPFPRGDSRGN
PSGGNGVQHRAGDQCPNPGCGGNQNFARTEBQCKAPKGFGLPFPFPGGRGR
GPGMGRGGRGGLMDRGPGMFRGGRGGRGDRGFGFGGRRGPGPGPGCG
LMEQNGRRGGRGGRGKMDKGEHBERDRPY"
1127. 1387
/molecule="RNA"
/bound_molecule="BWS"

```

```

polyA_signal      2162. .2167
                  /gene="EMS"
evidence=experimental
polyA_signal      2350. .2355
                  /gene="EMS"
evidence=experimental
polyA_site        2371
                  /gene="EMS"
ORIGIN

```

Query Match	99.9%	Score 2023.4	DB 9	length 2350
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2024, Conservative	0	Mismatches	1	Indels 0; Gaps 0;

QY	1	GACGGACGTTGAGAGAA	CAGAGAGAAAGAGAAAA	TGGCGTTCACGATTA	CAGTA	ACC	60				
Db	8	GACGGACGTTGAGAGAA	CAGAGAGAAAGAGAAAA	TGGCGTTCACGATTA	CAGTA	ACC	67				
QY	61	TATATGCCAAGCTG	CAGGCGCAGCAGGGCTTA	CAGTGTCTTA	CA	CCCCCGCAGCCCACTCA	AGA	120			
Db	68	TATATGCCAAGCTG	CAGGCGCAGCAGGGCTTA	CAGTGTCTTA	CA	CCCCCGCAGCCCACTCA	AGA	127			
QY	121	TATGACACAG	CAACCCAGGACATA	TGGGCA	CAAAAGCTAT	GAAACCTAT	TGGA	CAGCCCACT	180		
Db	128	TATGACACAG	CAACCCAGGACATA	TGGGCA	CAAAAGCTAT	GAAACCTAT	TGGA	CAGCCCACT	187		
QY	181	GATGTACGCTA	TATCCAGGCTCAG	ACCACTG	CAACCTAT	TGGGCA	CAACCCGCTTAT	GCACT	240		
Db	188	GATGTACGCTA	TATCCAGGCTCAG	ACCACTG	CAACCTAT	TGGGCA	CAACCCGCTTAT	GCACT	247		
QY	241	TCTTATGACAG	CTCTCCCACTGTGT	TATACCTAC	CTACCCCGCAGGAT	CAATCAG	CAG	300			
Db	248	TCTTATGACAG	CTCTCCCACTGTGT	TATACCTAC	CTACCCCGCAGGAT	CAATCAG	CAG	307			
QY	301	CCGTGTCAGGGGGT	TGSCACTGGTCTTAT	GATATCA	CAACCTGTGTA	CAGTCA	CAACCA	CACC	360		
Db	308	CCGTGTCAGGGGGT	TGSCACTGGTCTTAT	GATATCA	CAACCTGTGTA	CAGTCA	CAACCA	CACC	367		
QY	361	CAGGCTCTCTA	TGACAGTCA	GTCTGCAT	TATGGA	CTCAG	CTGTGT	TCCAG	CTTATGG	420	
Db	368	CAGGCTCTCTA	TGACAGTCA	GTCTGCAT	TATGGA	CTCAG	CTGTGT	TCCAG	CTTATGG	427	
QY	421	CAGCAGCCAG	AGAGCACTTGCA	CCCTACAA	AGCCGAGAT	TGGA	CAAGCCCA	CTGAG	ACT	480	
Db	428	CAGCAGCCAG	AGAGCACTTGCA	CCCTACAA	AGCCGAGAT	TGGA	CAAGCCCA	CTGAG	ACT	487	
QY	481	AGTCAACCTCA	ATCTAGCA	CAGGGGGTTA	CAAC	CAGAGCC	CAGCTAT	TGACAG	AGT	540	
Db	488	AGTCAACCTCA	ATCTAGCA	CAGGGGGTTA	CAAC	CAGAGCC	CAGCTAT	TGACAG	AGT	547	
QY	541	AACTAACAGTTA	TCCCAAGTA	CCTGGAG	CTACCCAT	GACAG	CACTG	CACTG	CACTTCA	600	
Db	548	AACTAACAGTTA	TCCCAAGTA	CCTGGAG	CTACCCAT	GACAG	CACTG	CACTG	CACTTCA	607	
QY	601	TCTTACCTCT	TACCAAGCTAT	TCTCTT	CAACAGCCG	CACTAGTTA	TGAT	TCAGAG	AGTAC	660	
Db	608	TCTTACCTCT	TACCAAGCTAT	TCTCTT	CAACAGCCG	CACTAGTTA	TGAT	TCAGAG	AGTAC	667	
QY	661	TCTCAGCAGAA	CACCTAT	TGGGCA	CCGAGCAG	CTAT	TGGA	CAGAGAG	TATGTC	720	
Db	668	TCTCAGCAGAA	CACCTAT	TGGGCA	CCGAGCAG	CTAT	TGGA	CAGAGAG	TATGTC	727	
QY	721	CAAAAGCAGT	ATGSGCAG	CAGCCTT	CCCACTAT	GTAT	ACCA	CCCCCA	ACTG	780	
Db	728	CAAAAGCAGT	ATGSGCAG	CAGCCTT	CCCACTAT	GTAT	ACCA	CCCCCA	ACTG	787	
QY	781	CAAGTCCAA	GTCAATAT	TAGCCAA	CAGAGCAG	CACTTAC	GGGCA	GAGAGTTC	ATCCGA	840	
Db	788	CAAGTCCAA	GTCAATAT	TAGCCAA	CAGAGCAG	CACTTAC	GGGCA	GAGAGTTC	ATCCGA	847	
QY	841	CAGAGCA	CCCCCAGT	AGATGGG	GTAT	TATGGG	AGAGG	CTG	TGAGAGATTTTTC	CGAGCA	900
Db	848	CAGAGCA	CCCCCAGT	AGATGGG	GTAT	TATGGG	AGAGG	CTT	TGAGAGATTTTTC	CGAGCA	907

QY 901 GGAGAGAACCGGAGCATGATGGCCCTGATTAACCGGGGCAAGGGAAGAGGGGATTTGAT 960  
DB 908 GGAGAGAACCGGAGCATGATGGCCCTGATTAACCGGGGCAAGGGAAGAGGGGATTTGAT 967  
QY 961 CGTGAAGCATGAGCAGAGGTGGGCGGGAGAGAGCGCGGTGATGAGCGCGCTGA 1020  
DB 968 CGTGAAGCATGAGCAGAGGTGGGCGGGAGAGAGCGCGGTGATGAGCGCGCTGA 1027  
QY 1021 GAGCGAGGTGGCTTCAATTAAGCTGTGAGCCATGATGAAGACCAATCTTGATCTA 1080  
DB 1028 GAGCGAGGTGGCTTCAATTAAGCTGTGAGCCATGATGAAGACCAATCTTGATCTA 1087  
QY 1081 GGCCCACTGTAATCCAGATGAAGACTCTGACACAGTGAATTTATGACAGATTA 1140  
DB 1088 GGCCCTCTCTGTAGATCCAGATGAAGACTCTGACACAGTGAATTTATGACAGATTA 1147  
QY 1141 AATGACAGTGAATCTGATGATCTGACAGACTTCTTTAAGCAGTGTGGGGTTGTTAG 1200  
DB 1148 AATGACAGTGAATCTGATGATCTGACAGACTTCTTTAAGCAGTGTGGGGTTGTTAG 1207  
QY 1201 ATGAACAGAGAACTGGGCAACCCATGATCACTTACCTGACAGAGAAACAGAAAG 1260  
DB 1208 ATGAACAGAGAACTGGGCAACCCATGATCACTTACCTGACAGAGAAACAGAAAG 1267  
QY 1261 CCCAAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGCAAGGCTGCCGTGGA 1320  
DB 1268 CCCAAAGGCGATGCCACAGTGTCTTATGAAGACCCACCTGCAAGGCTGCCGTGGA 1327  
QY 1321 TGGTTTGAATGGGAAGATTTTCAAGGGAGCAAACTTAAGTCTCCCTTGTCCGGAAGAG 1380  
DB 1328 TGGTTTGAATGGGAAGATTTTCAAGGGAGCAAACTTAAGTCTCCCTTGTCCGGAAGAG 1387  
QY 1381 CTTCCAAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGCAGAGCATGCCA 1440  
DB 1388 CTTCCAAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGCAGAGCATGCCA 1447  
QY 1441 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCTGTGGGGAGCCCATGGTCCGATGGA 1500  
DB 1448 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCTGTGGGGAGCCCATGGTCCGATGGA 1507  
QY 1501 GGCCTGAGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCGAGGGAGAC 1560  
DB 1508 GGCCTGAGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCGAGGGAGAC 1567  
QY 1561 CCCTTGAAGAGAGAAAGTCCAGACCGAGCTGAGAGCTGCAAGTGTCCCAATCCGGGT 1620  
DB 1568 CCCTTGAAGAGAGAAAGTCCAGACCGAGCTGAGAGCTGCAAGTGTCCCAATCCGGGT 1627  
QY 1621 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGCACAGTGTAAAGCCCAAGACCT 1680  
DB 1628 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGCACAGTGTAAAGCCCAAGACCT 1687  
QY 1681 GAAGGCTTCTCCCGACACCTTTCGCGCCCGGGGTGATGATGTCAGAGGTGAGCCCT 1740  
DB 1688 GAAGGCTTCTCCCGACACCTTTCGCGCCCGGGGTGATGATGTCAGAGGTGAGCCCT 1747  
QY 1741 GGTGAGCATGCGGAGAGAGAGGTGAGCTCATGATGATGATGATGATGATGATGATGAT 1800  
DB 1748 GGTGAGCATGCGGAGAGAGAGGTGAGCTCATGATGATGATGATGATGATGATGATGAT 1807  
QY 1801 AGAGGTGCGGTGTGAGAGCAAGGTGCTTCCGTGTGTGGCGGGGCAATGACCGAGGT 1860  
DB 1808 AGAGGTGCGGTGTGAGAGCAAGGTGCTTCCGTGTGTGGCGGGGCAATGACCGAGGT 1867  
QY 1861 GGTCTTGTGAGAGAGAGAGAGGTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
DB 1868 GGTCTTGTGAGAGAGAGAGAGAGGTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1927  
QY 1921 ATGAG 1980  
DB 1928 ATGAG 1987

QY 1981 CAGAGGCGCAGAGATGCGCCCTACTAGATGACAGAGCCCGCAGA 2025  
DB 1988 CAGAGGCGCAGAGATGCGCCCTACTAGATGACAGAGCCCGCAGA 2032

RESULT 4  
A36460  
LOCUS A36460  
DEFINITION Sequence 1 from Patent WO9323549.  
ACCESSION A36460  
VERSION A36460.1 GI:2293778  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
Aurias A., Delattre O., Desmaze C., Melot T., Peter M.,  
1 (bases 1 to 2371)  
AUTHORS  
TITLE  
JOURNAL  
NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE  
DEVELOPMENT OF CANCEROUS TUMORS  
Patent: WO 9323549-A 1 25-NOV-1993;  
CENTRE NAT RECH SCIEN (FR)  
COMMENT  
Other publication FR 2691475 931126  
Other publication JP 85009647 960206.  
FEATURES  
Location/Qualifiers  
1..2371  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/dev\_stage="FOETUS"

ORIGIN  
Query Match 99.3%; Score 2011.4; DB 6; Length 2371;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2012; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 GAGAACGAGAGAGAGAGAGAAATGCGGTCCAGGATTTACAGTACTATAGCCAAAGCT 72  
DB 1 GAGAACGAGAGAGAGAGAGAAATGCGGTCCAGGATTTACAGTACTATAGCCAAAGCT 60  
QY 73 GCAGGCCAGAGAGGCTACAGTGTCTTACACCGCCCAAGCCCTCAAGATTTGACAGAGC 132  
DB 61 GCAGGCCAGAGAGGCTACAGTGTCTTACACCGCCCAAGCCCTCAAGATTTGACAGAGC 120  
QY 133 ACCAGGCAATATGAGGCAACAAAGCTATGGAACCTATGAGACGCCACTGATGTCAGTAT 192  
DB 121 ACCAGGCAATATGAGGCAACAAAGCTATGGAACCTATGAGACGCCACTGATGTCAGTAT 180  
QY 193 ACCAGGCTCAGACCACTGCACTATGAGCAGACCGCTATGCAACTTCTTATGAGCAG 252  
DB 181 ACCAGGCTCAGACCACTGCACTATGAGCAGACCGCTATGCAACTTCTTATGAGCAG 240  
QY 253 CTTCCCACTGTGTTATCTACTGCACTGCCCCCAGGCAATACAGGAGCTGTCCAGGG 312  
DB 241 CTTCCCACTGTGTTATCTACTGCACTGCCCCCAGGCAATACAGGAGCTGTCCAGGG 300  
QY 313 TATGAGCACTGTGCTTATGATACCACTGCTGCTATGCAAGTCAACCAAGGCTCTCTAT 372  
DB 301 TATGAGCACTGTGCTTATGATATACCACTGCTGCTATGCAAGTCAACCAAGGCTCTCTAT 360  
QY 373 GCAGCTCACTGTGCAATATGAGCACTGAGCTGCTTATCAGGCTATGAGGAGCAGCAGCA 432  
DB 361 GCAGCTCACTGTGCAATATGAGCACTGAGCTGCTTATCAGGCTATGAGGAGCAGCAGCA 420  
QY 433 GCACTGTGCACTTACAGAGCCGAGAGATGGAACCAAGCCCACTGAGACTGTCAACTCA 492  
DB 421 GCACTGTGCACTTACAGAGCCGAGAGATGGAACCAAGCCCACTGAGACTGTCAACTCA 480  
QY 493 TCTAGCAGAGGGGGTTACAGACAGCCAGCTAGATATGAGCAGAGTAACTACAGTAT 552  
DB 481 TCTAGCAGAGGGGGTTACAGACAGCCAGCTAGATATGAGCAGAGTAACTACAGTAT 540

QY	553	CCCCAGGATACCTGGAGGTACCCCATGAGCAGCTGACCTGCATCTCACTCCCT	612
Db	541	CCCCAGGATACCTGGAGGTACCCCATGAGCAGCTGACCTGCATCTCACTCCCT	600
QY	613	ACGAGCTATTCCTTACACAGCCGACCTAGTTATGATCAGACAGTTACTCTCAGCAGAAC	672
Db	601	ACGAGCTATTCCTTACACAGCCGACCTAGTTATGATCAGACAGTTACTCTCAGCAGAAC	660
QY	673	ACCTATGGGCAACCGAGAGCTATGAGCAGAGAGTGCATATGGTCAACAAAGAGCTAT	732
Db	661	ACCTATGGGCAACCGAGAGCTATGAGCAGAGAGTGCATATGGTCAACAAAGAGCTAT	720
QY	733	GGGACGAGACGCTCCCACTAGTTAACCCACCCCAACTGGATCTTACAGCCAGCTCCAAGT	792
Db	721	GGGACGAGACGCTCCCACTAGTTAACCCACCCCAACTGGATCTTACAGCCAGCTCCAAGT	780
QY	793	CAATATGACCAACAGACGACGACTACGGGACGACAGATTATTCGACAGAACCAACCC	852
Db	781	CAATATGACCAACAGACGACGACTACGGGACGACAGATTATTCGACAGAACCAACCC	840
QY	853	AGTAGCATGGGTGTTTATGAGGACAGAGCTGAGAGATTTTCCGACACGAGAGAACCGG	912
Db	841	AGTAGCATGGGTGTTTATGAGGACAGAGCTGAGAGATTTTCCGACACGAGAGAACCGG	900
QY	913	AGCATGAGTGGCCCTGATTAACCGGCGGACAGGGGAAAGGGGGATTTGATCGTGAGGCATG	972
Db	901	AGCATGAGTGGCCCTGATTAACCGGCGGACAGGGGAAAGGGGGATTTGATCGTGAGGCATG	960
QY	973	AGCAGAGTGGCCGGGAGAGAGACGGCGTGGAAATGGCACAACGCTGAGACACGAGGTGGC	1032
Db	961	AGCAGAGTGGCCGGGAGAGAGACGGCGTGGAAATGGCACAACGCTGAGACACGAGGTGGC	1020
QY	1033	TTCAATTAAGCTGTGTGACCCCATGATGAAAGAACCAATCTTGAATCTTAGGCCACCTGTA	1092
Db	1021	TTCAATTAAGCTGTGTGACCCCATGATGAAAGAACCAATCTTGAATCTTAGGCCACCTGTA	1080
QY	1093	GATCCAGATGAAGACTCTGACAAACAGTGCATTTATGTACAAAGATTTAAATGACAGTGTG	1152
Db	1081	GATCCAGATGAAGACTCTGACAAACAGTGCATTTATGTACAAAGATTTAAATGACAGTGTG	1140
QY	1153	ACTCTAGATGATCTGGCAGACCTTCTTTAAGCAGTGTGGGGTTGGTAAAGATGAACAAGAA	1212
Db	1141	ACTCTAGATGATCTGGCAGACCTTCTTTAAGCAGTGTGGGGTTGGTAAAGATGAACAAGAA	1200
QY	1213	ACTGGGCAACCCATGATCAACATCTTACTTGAACAGAGAACGAGAACCCCAAGGCGAT	1272
Db	1201	ACTGGGCAACCCATGATCAACATCTTACTTGAACAGAGAACGAGAACCCCAAGGCGAT	1260
QY	1273	GCACAGTGTCTTATGAAGACCCACCTGCACATGCGCAGAGGTGCGGTGGAAATGGTTGATGGG	1332
Db	1261	GCACAGTGTCTTATGAAGACCCACCTGCACATGCGCAGAGGTGCGGTGGAAATGGTTGATGGG	1320
QY	1333	AAAGATTTTCAAGGGAGCAAACTTAAAGTCTTCCCTTGTCTGGAAGAAAGCCTTCAATGAC	1392
Db	1321	AAAGATTTTCAAGGGAGCAAACTTAAAGTCTTCCCTTGTCTGGAAGAAAGCCTTCAATGAC	1380
QY	1393	AGTATGCGGGGTGTGTCTGCAACCCCGTGAAGGGCAGAGGCATGCAACCACTCCGTGGA	1452
Db	1381	AGTATGCGGGGTGTGTCTGCAACCCCGTGAAGGGCAGAGGCATGCAACCACTCCGTGGA	1440
QY	1453	GGTCCAGAGGCCCCAGAGAGTCTGTGGGGGACCCATGGGTGTGATGGAGGCGCGTGAAGGA	1512
Db	1441	GGTCCAGAGGCCCCAGAGAGTCTGTGGGGGACCCATGGGTGTGATGGAGGCGCGTGAAGGA	1500
QY	1513	GATAGAGGAGGCTTCCCTTCAAGAGAGACCCCGGGGTTTCCGAGGGAAACCCCTTGAAGGA	1572
Db	1501	GATAGAGGAGGCTTCCCTTCAAGAGAGACCCCGGGGTTTCCGAGGGAAACCCCTTGAAGGA	1560
QY	1573	GGAAACGTCCAGACCCGAGCTGAGAACCTGGCAGTGTCCCAATCCGGGTGTGAGAAACAG	1632
Db	1561	GGAAACGTCCAGACCCGAGCTGAGAACCTGGCAGTGTCCCAATCCGGGTGTGAGAAACAG	1620

QY	1633	AACTTGCCCTGGAGAA	CAGAGTGC	AAACAGTGTAAAGGCC	CCAAAGCCTGAAGCTTCCTC	1652	
Db	1621	AACTTCGCCCTGGAGAA	CAGAGTGC	AAACAGTGTAAAGGCC	CCAAAGCCTGAAGCTTCCTC	1680	
QY	1693	CCGCACCTCTTCCGCC	CCCGGGGTGTGAT	CGTGGC	CAGAGTGGCCCTTGTTGGCATCCGG	1752	
Db	1681	CCGCACCTCTTCCGCC	CCCGGGGTGTGAT	CGTGGC	CAGAGTGGCCCTTGTTGGCATCCGG	1740	
QY	1753	GGAGGAAAGAGTGGCC	CTCATGAA	TCGTGGTGTGTC	CCCGGTGGAAATGTT	CAGAGTGGCCGT	1812
Db	1741	GGAGGAAAGAGTGGCC	CTCATGAA	TCGTGGTGTGTC	CCCGGTGGAAATGTT	CAGAGTGGCCGT	1800
QY	1813	GGTGGAGA	CAGAGTGGCTTC	CCGTTGGTGGCC	CGGGGACATGGA	CCGAGGTGGCTTGTGTGTA	1872
Db	1801	GGTGGAGA	CAGAGTGGCTTC	CCGTTGGTGGCC	CGGGGACATGGA	CCGAGGTGGCTTGTGTGTA	1860
QY	1873	GGAAGACGAGGTGGC	CCCTGGGGGGGCC	CCCTGGAC	CTTTGTATGGA	CAGATGGAGNA	1932
Db	1861	GGAAGACGAGGTGGC	CCCTGGGGGGGCC	CCCTGGAC	CTTTGTATGGA	CAGATGGAGNA	1920
QY	1933	AGAGAGAGA	CGTGGAGGA	CCCTGGAAAAT	TGATPAAAGGCAC	GCCTCAGAGACCGAGA	1992
Db	1921	AGAGAGAGA	CGTGGAGGA	CCCTGGAAAAT	TGATPAAAGGCAC	GCCTCAGAGACCGAGA	1980
QY	1993	GATCGGCCCTTACTGAT	GCAGAGCA	CCCCGCAGA	2025		
Db	1981	GATCGGCCCTTACTGAT	GCAGAGCA	CCCCGCAGA	2013		

RESULT 5	AR080100	2371 bp	DNA	linear	PAT 31-AUG-2000
LOCUS	Sequence 1 from patent US 5968734.				
DEFINITION	AR080100				
ACCESSION	AR080100.1	GI:10006835			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 2371)				
	Auriaz,A., Delattre,O., Desmazes,C., Melot,T., Peter,M.,				
	Plougas,et al., B., Thomas,G. and Zucman,J.				
TITLE	Nucleic acid corresponding to a gene of chromosome 22 involved in recurrent chromosomal translocations associated with the development of cancerous tumors, and nucleic acids of fusion resulting from said translocations				
JOURNAL	Patent: US 5968734-A 119-Oct-1999;				
FEATURES	Location/Qualifiers				
Source	1..2371				

**ORIGIN**

Query Match	99.3%	Score 2011.4	DB 6	length 2371
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2012, Conservative	0	Mismatches	1	Indels 0; Gaps 0;

QY	13	GAGAACGAGGAGGAGGAGAAATATGGCCCTCAGCATTCACCTATATAGCCAAAGCT	72
Db	1	GAGAACGAGGAGGAGGAGAAATATGGCCCTCAGCATTCACCTATATAGCCAAAGCT	60
QY	73	GCAGCGCAGCAGGGCTACAGTGCCTTACACCGCCAGCCCACTCAAGATATGACACAGCC	132
Db	61	GCAGCGCAGCAGGGCTACAGTGCCTTACACCGCCAGCCCACTCAAGATATGACACAGACC	120
QY	133	ACCCAGGCAATATGGGCAACAAAGCTATATGAACTTATGACAGGCCCACTGATGTCAAGCTAT	192
Db	121	ACCCAGGCAATATGGGCAACAAAGCTATATGAACTTATGACAGGCCCACTGATGTCAAGCTAT	180
QY	193	ACCCAGGCTCAGACCACTGCAACCTATATGGGACAGCCGCTATGCAACTTCTTATGACAG	252
Db	181	ACCCAGGCTCAGACCACTGCAACCTATATGGGACAGCCGCTATGCAACTTCTTATGACAG	240

QY 253 CCTCCACGTGGTATATCTACTCAACTGACCCGCCAGGCAATACAGCAGCCCTGTCACAGAGG 312  
 Db 241 CTTCCCACTGGTATATCTACTCAACTGACCCGCCAGGCAATACAGCAGCCCTGTCACAGAGG 300  
 QY 313 TATGGCACTGGTGTATATGATACCACTGCTACAGTCAACACACCCAGGCTCTCTAT 372  
 Db 301 TATGGCACTGGTGTATATGATACCACTGCTACAGTCAACACACCCAGGCTCTCTAT 360  
 QY 373 GCAAGCTAGTCTGATATGCACTGACCTGCTTATCCAGCTATGAGGACAGCCAGCA 432  
 Db 361 GCAAGCTAGTCTGATATGCACTGACCTGCTTATCCAGCTATGAGGACAGCCAGCA 420  
 QY 433 GCCACTGCACTTACCAAGCCGAGAGTGAACCAAGCCCACTGAGCTAAGTCAACTCA 492  
 Db 421 GCCACTGCACTTACCAAGCCGAGAGTGAACCAAGCCCACTGAGCTAAGTCAACTCA 480  
 QY 493 TCTAGCAGAGGGGGTTTCAACCAAGCCCACTGAGTATGGAAGAGTAACTACAGTTAT 552  
 Db 481 TCTAGCAGAGGGGGTTTCAACCAAGCCCACTGAGTATGGAAGAGTAACTACAGTTAT 540  
 QY 553 CCCCAAGTACTGGAGAGTACCCCAATGACCCAGTCACTGCACTCTCACTCTCTCT 612  
 Db 541 CCCCAAGTACTGGAGAGTACCCCAATGACCCAGTCACTGCACTCTCTCACTCTCTCT 600  
 QY 613 ACCAGCTATCTCTCTACACAGCCGACTAGTTATGATGAGAGAGTAACTCTCAGCAGAAC 672  
 Db 601 ACCAGCTATCTCTCTACACAGCCGACTAGTTATGATGAGAGAGTAACTCTCAGCAGAAC 660  
 QY 673 ACCTATGGGCAACCGAGCAGCTATGGAAGAGAGTATGATGATGATGATGATGATGAT 732  
 Db 661 ACCTATGGGCAACCGAGCAGCTATGGAAGAGAGTATGATGATGATGATGATGATGAT 720  
 QY 733 GGGCAGCAGCTCTCCCACTAGTATCCCACTGCTCAAGCTCTCAAGCTCTCAAGCTCT 792  
 Db 721 GGGCAGCAGCTCTCCCACTAGTATCCCACTGCTCAAGCTCTCAAGCTCTCAAGCTCT 780  
 QY 793 CAATATGAGCAAGAGCAGCAGCTAGGGCAGAGAGTCAATTCGAGAGAGAGAGAGAGAG 852  
 Db 781 CAATATGAGCAAGAGCAGCAGCTAGGGCAGAGAGTCAATTCGAGAGAGAGAGAGAGAG 840  
 QY 853 AGTAGCAGTGGGTGTTATGAGGAGAGAGTCTGAGAGATTTTCCGAGACGAGAGAGAGAG 912  
 Db 841 AGTAGCAGTGGGTGTTATGAGGAGAGAGTCTGAGAGATTTTCCGAGACGAGAGAGAGAG 900  
 QY 913 AGCATAGTGGCCCTGATTAACCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972  
 Db 901 AGCATAGTGGCCCTGATTAACCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 973 AGCAGAGTGGGGGAG 1032  
 Db 961 AGCAGAGTGGGGGAG 1020  
 QY 1033 TTCAATAGGCTGGTGGACCCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1092  
 Db 1021 TTCAATAGGCTGGTGGACCCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 QY 1093 GATCCGATGAG 1152  
 Db 1081 GATCCGATGAG 1140  
 QY 1153 ACTCTGATGATGATGAG 1212  
 Db 1141 ACTCTGATGATGATGAG 1200  
 QY 1213 ACTGGGCAACCCATGATGAG 1272  
 Db 1201 ACTGGGCAACCCATGATGAG 1260  
 QY 1273 GCCACAGTGTCTATGAG 1332  
 Db 1261 GCCACAGTGTCTATGAG 1320  
 QY 1333 AAAGATTTTCAAGGAG 1392

Db 1321 AAAGATTTTCAAGGAG 1380  
 QY 1393 AGTATGCGGGGTGGTCTGCCACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452  
 Db 1381 AGTATGCGGGGTGGTCTGCCACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1453 GGTTCAG 1512  
 Db 1441 GGTTCAG 1500  
 QY 1513 GATAG 1572  
 Db 1501 GATAG 1560  
 QY 1573 GAAAGAGTCCAG 1632  
 Db 1561 GAAAGAGTCCAG 1620  
 QY 1633 AACTTGGCTGAG 1692  
 Db 1621 AACTTGGCTGAG 1680  
 QY 1693 CCGCAGACCTTTCCGAG 1752  
 Db 1681 CCGCAGACCTTTCCGAG 1740  
 QY 1753 GAG 1812  
 Db 1741 GAG 1800  
 QY 1813 GGTGAG 1872  
 Db 1801 GGTGAG 1860  
 QY 1873 GGAAG 1932  
 Db 1861 GGAAG 1920  
 QY 1933 AG 1992  
 Db 1921 AG 1980  
 QY 1993 GATGCGCCTTACTAGATGAG 2025  
 Db 1981 GATGCGCCTTACTAGATGAG 2013

RESULT 6  
 BC004817  
 LOCUS 2182 bp mRNA 11near PRI 29-JUN-2004  
 DEFINITION Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 EMS, mRNA (cDNA clone MGC:5432 IMAGE:449145), complete cds.  
 ACCESSION BC004817  
 VERSION BC004817.1 GI:13435562  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 2182)  
 Strausberg, R.L., Fellgoid, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Zeeberg, B., Burow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diachenko, L., Martusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Uedl, T.B., Toshitani, S.,  
 Carinici, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J., Gunaratne, P.H., Richards, S.,  
 Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

JOURNAL

REMARK

COMMENT

Phley, J., Halton, E., Kelleman, M., Madan, A.C., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schercher, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2182)

Strauberg, R.

Direct Submission

Submitted (21-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [gagabs-remail.nih.gov](mailto:gagabs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amgdbcm.tmc.edu](mailto:amgdbcm.tmc.edu)

Gunnarsson, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loueseg, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanaev, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>

Series: IRAX Plate: 3 Row: 1 Column: 6

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers

1..2182

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:5432 IMAGE:3449145"

/tissue\_type="Placenta, chorioarctoma"

/clone\_id="NH\_MGC\_10"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..2182

/gene="EWSR1"

/note="Synonym: EWS"

/db\_xref="LOCUSID:2130"

/db\_xref="MIM:133450"

20..1990

/gene="EWSR1"

/codon\_start=1

/product="Ewing sarcoma breakpoint region 1, isoform EWS"

/protein\_id="AAH04817.1"

/db\_xref="GI:13435963"

/db\_xref="LOCUSID:2130"

/db\_xref="MIM:133450"

/translation="MASTSDTSTYQOAAAGGYSAVTAPTOGTACTOATGGSYTYTTPDVSATOAATTATVGCTATVATSGPPGYTTTAPAPOAVSQPVGGYGVGAYDTITTAATTTTGASVAAGYATQPAVPAYGQPATPATPTPDGAKPRTSOPOSTGVNQPSLXGOSNVASYGPVPGSPYMPQATPASPYSPTSSNOTPSYOASSYDOONTYGSSSVYGOSSSYGOOSSYGOCPPTSYPOTGSYGAAPYGOSSSYGOSSSFODHSSMGVYGOSSGFSGEGENRSMSGPDNRGRGRGRGDROGMBSRGGRGCGMGSAGEFGTRKPGMBEGPLDLGPVDPEDSDNSAIYVGLNDSTVLDDLDADFRQCGVAVWNRKTCGPMIHTLYDKETGPKGDAIVSEDPITKAAVIEWFDGDFGSKLTSLAKKPPMNSRGLIPPREGEMPEPLRGGPCGPGCGPMGMRGDRGDRGPPPRGSRBNPSSGNNVORHAGDMQCNPNGGNONFAWTCECNCKAPKEGFLPPEPPPGDRGGRGPGRCGRGRGLMDRGRCGMFRGCRGDRGRCGRCGMDBRGRCGGRGRCGPPRPMLEQNGRRGRGGRGKMDKSHNRERRDRFY"

[illegible]

QY	1038	TAAGCTGTGGACCCCAAGGATGAAGAGCAAGATCTTGAATCTAAGCCCACTGTAGATCC	1097
Db	1021	TAAGCTGTGGACCCCAAGGATGAAGAGCAAGATCTTGAATCTAAGCCCACTGTAGATCC	1080
QY	1098	AGATGAAGACTCTGACAAAGGTCAATTTATGTACAGAGATTAAATGACAGTGTGACTCT	1157
Db	1081	AGATGAAGACTCTGACAAAGGTCAATTTATGTACAGAGATTAAATGACAGTGTGACTCT	1140
QY	1158	AGATGATCTGGAGACTCTTTAGCAGTGTGGGTGTGTAAGATGAACAAGAGACTGG	1217
Db	1141	AGATGATCTGGAGACTCTTTAGCAGTGTGGGTGTGTAAGATGAACAAGAGACTGG	1200
QY	1218	GCAACCCATGATCCCATCTACCTGTGACAAAGGAAACAGAAAAGCCCAAGGCGATGCCAC	1277
Db	1201	GCAACCCATGATCCCATCTACCTGTGACAAAGGAAACAGAAAAGCCCAAGGCGATGCCAC	1260
QY	1278	AGTGTCTATGAGAACCCACCACTGCCAAGGCTGCGTGTGANTGGTTTGAATGGAGAA	1337
Db	1261	AGTGTCTATGAGAACCCACCACTGCCAAGGCTGCGTGTGANTGGTTTGAATGGAGAA	1320
QY	1338	TTTTCAAGGAGCAAACTTTAAAGTCTCCCTTGTCTGGAAAGAAAGCTCCAAATGACAGTAT	1397
Db	1321	TTTTCAAGGAGCAAACTTTAAAGTCTCCCTTGTCTGGAAAGAAAGCTCCAAATGACAGTAT	1380
QY	1398	GCGGGGTGTGTCTGACACCCCGTGAAGGAGAGGACATGACACACCACTCCGATGAGCTCC	1457
Db	1381	GCGGGGTGTGTCTGACACCCCGTGAAGGAGAGGACATGACACACCACTCCGATGAGCTCC	1440
QY	1458	AGAGAGCCCAAGAGAGTCTCTGGGGAGCCCATGGGTGCAATGGAGGCGTGTGAGAGATAG	1517
Db	1441	AGAGAGCCCAAGAGAGTCTCTGGGGAGCCCATGGGTGCAATGGAGGCGTGTGAGAGATAG	1500
QY	1518	AGGAGGCTTTCCTCCAAAGAGACCCCCGGGGTTCCCGAAGAAACCTCTTGGAGAGAGAA	1577
Db	1501	AGGAGGCTTTCCTCCAAAGAGACCCCCGGGGTTCCCGAAGAAACCTCTTGGAGAGAGAA	1560
QY	1578	CGTCCAGACCGAGAGCTGTGAGACTGTGCCAAATCCGGGTGTGTGAAAACAGAACTT	1637
Db	1561	CGTCCAGACCGAGAGCTGTGAGACTGTGCCAAATCCGGGTGTGTGAAAACAGAACTT	1620
QY	1638	CGCTTGAAGAACAGAGTGCACAGATGTAAAGGCCCAAGGCTTGAGAGCTTCTCCGCGC	1697
Db	1621	CGCTTGAAGAACAGAGTGCACAGATGTAAAGGCCCAAGGCTTGAGAGCTTCTCCGCGC	1680
QY	1698	AACCTTTCGGCCCCCGGGTGTGTGATCTGTGTGACAGAGTGGCCCCCTGTGTGCATTCGCGGAGG	1757
Db	1681	AACCTTTCGGCCCCCGGGTGTGTGATCTGTGTGACAGAGTGGCCCCCTGTGTGCATTCGCGGAGG	1740
QY	1758	AAGAGTGTGGCTCATGATTCGTGTGTGATCCCGGTGGAATGTTTCAAGAGTGTGCGGTGTGG	1817
Db	1741	AAGAGTGTGGCTCATGATTCGTGTGTGATCCCGGTGGAATGTTTCAAGAGTGTGCGGTGTGG	1800
QY	1818	AGACAGAGTGTGGCTTCCGTGTGTGCGCGGGGCAATGACCCAGAGTGTGCTTGTGTGAGAGAG	1877
Db	1801	AGACAGAGTGTGGCTTCCGTGTGTGCGCGGGGCAATGACCCAGAGTGTGCTTGTGTGAGAGAG	1860
QY	1878	AACGAGTGTGGCCCTGTGGGGGGCCCCCTGTGACCTTTGATGGAACAGATGTGGAAGAAACAGAG	1937
Db	1861	AACGAGTGTGGCCCTGTGGGGGGCCCCCTGTGACCTTTGATGGAACAGATGTGGAAGAAACAGAG	1920
QY	1938	AGGAGCTGTGAGGACTGTGAAAAAATGTGATTAAGGCGAGACCGTCCAGAGCGCAGAGATCG	1997
Db	1921	AGGAGCTGTGAGGACTGTGAAAAAATGTGATTAAGGCGAGACCGTCCAGAGCGCAGAGATCG	1980
QY	1998	GCCCTACTAGATGACAGAGACCCCGCAGA	2025
Db	1981	GCCCTACTAGATGACAGAGACCCCGCAGA	2008

RESULT 7  
CR456490  
2189 bp  
mRNA  
linear  
PRI 25-MAY-2004

DEFINITION					
Homo sapiens EWSR1 full length open reading frame (ORF)	cDNA clone				
ACCESSION					
CR456490.1	GI:47678510				
VERSION					
CDNA; chromosome 22; ORF.					
KEYWORDS					
SOURCE					
ORGANISM					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE					
1 (baaes 1 to 2189)					
Collins,J.E., Wright,C.L., Edwards,C.A., Davis,M.P., Grinham,J.A.,					
Cole,C.G., Goward,M.E., Aguado,B., Mallya,M., Mokrab,Y.,					
Huckle,E.J., Beare,D.M. and Dunham,I.					
TITLE					
Direct Submission					
JOURNAL					
Submitted (24-MAY-2004) Sanger Institute, Hinxton, Cambridgeshire,					
CB10 1SA, UK. E-mail enquiries: c22@sanger.ac.uk Manuscript					
COMMENT					
Sanger Institute name : pGEM_EWSR1					
Homo sapiens CDNA sequence. This sequence was generated as part of					
The Wellcome Trust Sanger Institute program to isolate cDNA clones					
representing the full length open reading frame of well annotated					
protein coding genes on human chromosome 22. For more information					
see http://www.sanger.ac.uk/HGP/Chr22/.					
FEATURES					
Source					
1..2189					
/organism="Homo sapiens"					
/mol_type="mRNA"					
/db_xref="taxon:9606"					
/chromosome="22"					
/clone="pGEM_EWSR1"					
/lab_host="UM109"					
1..2189					
/gene="EWSR1"					
34..2001					
/gene="EWSR1"					
/codon_start=1					
/protein_id="CA930376.1"					
/db_xref="gi:47678511"					
/translation="MASTDYSTYSQAAAGQSYSAVTAOPTGVATQTQAVQSGYGY					
GQPDVSYTQAQTATYGQTAIYAISYSCQPTGYTTPAPQASQPVQSYGTGADITTT					
ATVTTOASVAAQSAYGTOPAYPAVGQAPATAATPRDGNKPYETSSPOSSTGVNO					
PSLGGGSNSYSPOVPGSTYMPALTAPSPYPTSYSTPSTPSYDSSSOONTYGQPS					
SYGOSSSYGOSSSYGOQPPTSPYPTQTSYQAQBSQYSQSSSYGOQSSFPDDHSNMG					
VYGESGSGFGSENRSMGSDGNKRGRGGRDRGMSRGRRGGGGMGAEERGFKTK					
PGGMDGEGPLDLIDPPVDLPEDSDNSA1YYGLNDSTVLDDLDFPKQCVGVNNKKTT					
GQPMIH1LYDKETKGPKGDAIVSYEDPIAKAAVEMWDGDFOSSKLKVALARKPEPM					
NMSGAGLPREBRGMPPLRLGGPGCGPGGPMGRMGRGSRGSGFPFRGSRGSRNP					
SGGNGVOHRPADGMCNPGCGNQNFARTCSNCKAKPKRGAFLPPRPFGGGRGREGS					
PGNRGGRGGLMDRGCPGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG					
MEQGGRRGGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR					
ORIGIN					
Query Match	99.1%;	Score 2006;	DB 9;	Length 2189;	
Best Local Similarity	99.9%;	Pred. No. 0;			
Matches 2019;	Conservative 0;	Mismatches 0;	Indels 3;	Gaps 1;	
Db	1	GGAAGTTGAGAAGAACGAGAGAGAGAGAAAATGCGCTTCACGATTACAGTACTAT	63		
Oy	4	GGAAGTTGAGAAGAACGAGAGAGAGAGAAAATGCGCTTCACGATTACAGTACTAT	63		
Db	1	GGAAGTTGAGAAGAACGAGAGAGAGAGAAAATGCGCTTCACGATTACAGTACTAT	60		
Oy	64	AGCCAGCTGCAGCGCAGCAGCGGCTTACAAGTCTTACACCGGCCAGGCCACTCAAAGATAT	123		
Db	61	AGCCAGCTGCAGCGCAGCAGCGGCTTACAAGTCTTACACCGGCCAGGCCACTCAAAGATAT	120		
Oy	124	GCACGACCACCCAGGGCATATGCGCAACAAAGCTATGGAACCTATTGACAGGCCCATGAT	183		
Db	121	GCACGACCACCCAGGGCATATGCGCAACAAAGCTATGGAACCTATTGACAGGCCCATGAT	180		
Oy	184	GTCAGCTATACCAAGGCTCAGACACCTGCAACTATGGGAGACCGGCTTAGCACTTCT	243		
Db	181	GTCAGCTATACCAAGGCTCAGACACCTGCAACTATGGGAGACCGGCTTAGCACTTCT	240		
Oy	244	TATGACAGCTCCCACTGTTATATCTTCAACTGCCCCCAGGAGATACAGCACCTT	303		



Db	241	TATGAGCAGCCTCCCACTGGTTATATCTACAACTGCCCCAGAGATACAGCAGCCT	300
Qy	304	GTCCAGGGGGTATGGCACTGTGGCTTTATGATATACCACTGCTTACATGTCACCAACCAG	363
Db	301	GTCCAGGGGGTATGGCACTGTGGCTTTATGATATACCACTGCTTACATGTCACCAACCAG	360
Qy	364	GCCTCTTATGAGCTCAAGCTCTGCATATGAGGACCTAGCCTGCTTATCCAGCCTATGAGGAC	423
Db	361	GCCTCTTATGAGCTCAAGCTCTGCATATGAGGACCTAGCCTGCTTATCCAGCCTATGAGGAC	420
Qy	424	CAGCCAGAGCCACTGTCACCTTACAAAGCCGAGATGMAAACGCCCACTGACCTAGT	483
Db	421	CAGCCAGAGCCACTGTCACCTTACAAAGCCGAGATGMAAACGCCCACTGACCTAGT	480
Qy	484	CAACTCAATCTTAGCACAAGGGGGTTTACAACGCCCACTGACCTTAGATATGACAAGATAC	543
Db	481	CAACTCAATCTTAGCACAAGGGGGTTTACAACGCCCACTGACCTTAGATATGACAAGATAC	540
Qy	544	TACAGTTATCCCAAGATACCTGGGAGCTAACCCATGAGCCAGATCACTGCACCTCAATCC	603
Db	541	TACAGTTATCCCAAGATACCTGGGAGCTAACCCATGAGCCAGATCACTGCACCTCAATCC	600
Qy	604	TACCTCTCAAGCTATTTCTCTTACAGAGCCGACTAGTTATGATCAGAGCAGTTACTCT	663
Db	601	TACCTCTCAAGCTATTTCTCTTACAGAGCCGACTAGTTATGATCAGAGCAGTTACTCT	660
Qy	664	CAGCAGAACACTATGAGGCAACCGAGCAGCTATGGAACAGCAGAGTAGTACTATGTGCACAA	723
Db	661	CAGCAGAACACTATGAGGCAACCGAGCAGCTATGGAACAGCAGAGTAGTACTATGTGCACAA	720
Qy	724	AGCAGCTATGGGACACAGCCTCCCACTAGTTAACCAACCCCAACTGATCCTACAGCCAA	783
Db	721	AGCAGCTATGGGACACAGCCTCCCACTAGTTAACCAACCCCAACTGATCCTACAGCCAA	780
Qy	784	GCTCCAACTCATATATAGCCAAACAGAGCAGAGCTCAGGGCAGCAGAGTTCAATCCGACAG	843
Db	781	GCTCCAACTCATATATAGCCAAACAGAGCAGAGCTCAGGGCAGCAGAGTTCAATCCGACAG	840
Qy	844	GACCAACCCCACTAGAGATGGGTGTTTATGAGGACAGAGTCTGAGAGATTTCCGGACACAGA	903
Db	841	GACCAACCCCACTAGAGATGGGTGTTTATGAGGACAGAGTCTGAGAGATTTCCGGACACAGA	900
Qy	904	GAGAACCCGAGCATGATGATGGCCCTGATATAACCGGGCACAAGGGAGATTGATGCT	963
Db	901	GAGAACCCGAGCATGATGATGGCCCTGATATAACCGGGCACAAGGGAGATTGATGCT	960
Qy	964	GGAAGCATGAGCAGAGGTGGGCGGGGAGGAGAACCGGCTGGAATGGGACAGCGCTGAGAG	1023
Db	961	GGAAGCATGAGCAGAGGTGGGCGGGGAGGAGAACCGGCTGGAATGGGACAGCGCTGAGAG	1017
Qy	1024	CGAGGTGGCTTCAATATAGCCTGATGGAACCCATGATGTAAGAACCCAGATCTTGATCTAGGC	1083
Db	1018	CGAGGTGGCTTCAATATAGCCTGATGGAACCCATGATGTAAGAACCCAGATCTTGATCTAGGC	1077
Qy	1084	CAACTGTATGATCCAGATGTAACACTCTGACAACTGTCAACTTATATGTAACAAGATTAAT	1143
Db	1078	CAACTGTATGATCCAGATGTAACACTCTGACAACTGTCAACTTATATGTAACAAGATTAAT	1137
Qy	1144	GACAGTGTACTCTAGATGATCTGGCAGACTCTTTTAAGCACTGTGGGGTGTGTAAGTG	1203
Db	1138	GACAGTGTACTCTAGATGATCTGGCAGACTCTTTTAAGCACTGTGGGGTGTGTAAGTG	1197
Qy	1204	AACAAGAGAACTGGGCAACCCATGATCCACATCTACCTGGAACAGAGAAAGCC	1263
Db	1198	AACAAGAGAACTGGGCAACCCATGATCCACATCTACCTGGAACAGAGAAAGCC	1257
Qy	1264	AAAGGCGATGCCAAGTGTCTTATGAAGCCCAACCACTGGCAAGGCTGCGGTGAATGG	1332
Db	1258	AAAGGCGATGCCAAGTGTCTTATGAAGCCCAACCACTGGCAAGGCTGCGGTGAATGG	1317
Qy	1324	TTTGTATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCGAAGAAAGCT	1383
Db	1318	TTTGTATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCGAAGAAAGCT	1377

QY	1381	CCAAATGAACAAGTATGCGGGGTGGCTTCGCAACCCCGTAGGGGACAGGGCATGCAACACCA	1443
Db	1378	CCAAATGAACAAGTATGCGGGGTGGCTTCGCAACCCCGTAGGGGACAGGGCATGCAACACCA	1437
QY	1444	CTCCGTGAGAGTCCAGAGAGGCCCAAGAGGTCCTTGGGGAACTTAGGGTCCGATGGAGGC	1503
Db	1438	CTCCGTGAGAGTCCAGAGAGGCCCAAGAGGTCCTTGGGGAACTTAGGGTCCGATGGAGGC	1497
QY	1504	CGTAGAGAGATGAGAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCGAGGGAAACCC	1563
Db	1498	CGTAGAGAGATGAGAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCGAGGGAAACCC	1557
QY	1564	TCTGAGAGAGAAACGTCCAGCAACCGAGCTTGAGACTGCGATGTCCCAATCCGGGTTGT	1623
Db	1558	TCTGAGAGAGAGAAACGTCCAGCAACCGAGCTTGAGACTGCGAGTGTCCCAATCCGGGTTGT	1617
QY	1624	GGAAGACCAAACTTCGCTTGAGAGAACAGAGTGCACCAAGTGTAAAGGCCCCCAAGACTGAA	1683
Db	1618	GGAAGACCAAACTTCGCTTGAGAGAACAGAGTGCACCAAGTGTAAAGGCCCCCAAGACTGAA	1677
QY	1684	GGCTTCTCCCGCACCCCTTCCGCCCCCGGGGTGTGATCGTGGCAGAGGTGCGCTTGT	1743
Db	1678	GGCTTCTCCCGCACCCCTTCCGCCCCCGGGGTGTGATCGTGGCAGAGGTGCGCTTGT	1737
QY	1744	GGCATGCGGGAGGAGAGAGGTGGCTCATGAGATGTTGTGTCCCGTGGATTTTCA	1803
Db	1738	GGCATGCGGGAGGAGAGAGGTGGCTCATGAGATGTTGTGTCCCGTGGAGATTTCA	1797
QY	1804	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCCTGTGTGGCCGGGGCATGGAACGAGGTGGC	1863
Db	1798	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCCTGTGTGGCCGGGGCATGGAACGAGGTGGC	1857
QY	1864	TTTGTGTGAGAGAACAGAGTGGCCCTTGGGGGGGCCCCCTGACCTTTTGATGAAACAATG	1923
Db	1858	TTTGTGTGAGAGAACAGAGTGGCCCTTGGGGGGGCCCCCTGACCTTTTGATGAAACAATG	1917
QY	1924	GGAGAGAGAGAGAGAGACCTTGAGAGACCTTGAGAAATGATTAAGCGAGACACCTGTAG	1983
Db	1918	GGAGAGAGAGAGAGAGACCTTGAGAGACCTTGAGAAATGATTAAGCGAGACACCTGTAG	1977
QY	1984	GAGGCGAGAGATGGAGCCCTACTAGATGACAGACCCCGCAGA	2025
Db	1978	GAGGCGAGAGATGGAGCCCTACTAGATGACAGACCCCGCAGA	2019

	RESULT	8		
	BC072442			
	LOCUS			
	DEFINITION			
	ACCESSION			
	VERSION			
	KEYWORDS			
	SOURCE			
	ORGANISM			
REFERENCE				
AUTHORS				

  

BC072442	2164 bp	mRNA	linear	PRI 30-JUN-2004
Homo sapiens Ewing sarcoma breakpoint region 1,				transcript variant
EMS, mRNA (cDNA clone MGC:87992 IMAGE:6047805),				complete cds.
BC072442				
BC072442.1	GI:48734726			
MGC:				
Homo sapiens (human)				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate;				Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homini-				didae; Homo.
1 (bases 1 to 2164)				
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,				Klausner R.D., Collins,F.S., Wagner,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhattacharya,N.K.,				Hopkins,R.F., Jordan,H., Moore,T., Mak,S.I., Wang,J.Y., Hsieh,P.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,				Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.T.,
Scheer,T.E., Brownstein,M.J., Udell,T.B., Toshiba,Y.,				Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.U., Bosak,S.A., McEwan,P.J.,				McKernan,K.U., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huily,S.W.,				Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Falley,U., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,				Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,





Db 961 GAGAGACGCGGTGGAAATG---GCGCTGAGAGAGAGGTGGCTTCAATAGCCTGGTGA 1017  
 QY 1051 CCCATGATGAGAGACCAAGATCTTTGATCTAGGCCACCTGTATGATCAAGTAAGACTT 1110  
 Db 1018 CCCATGATGAGAGACCAAGATCTTTGATCTAGGCCACCTGTATGATCAAGTAAGACTT 1077  
 QY 1111 GACAAGAGTCAATTTATGTAACAAGATTAAAGATGATGATCTAGATGATCTAGGCA 1170  
 Db 1078 GACAAGAGTCAATTTATGTAACAAGATTAAAGATGATGATCTAGATGATCTAGGCA 1137  
 QY 1171 GACTTCTTTAAGAGATGAGGTGGTGTGTAAGATGTAACAAGAACTGGGCAACCATGATC 1230  
 Db 1138 GACTTCTTTAAGAGATGAGGTGGTGTGTAAGATGTAACAAGAACTGGGCAACCATGATC 1197  
 QY 1231 CACATCTACTGAGCAAGAAACAGAAAGCCCAAGGCGATGCGACAGTGTCTTATGAA 1290  
 Db 1198 CACATCTACTGAGCAAGAAACAGAAAGCCCAAGGCGATGCGACAGTGTCTTATGAA 1257  
 QY 1291 GACCCACCCCACTGCGAAGGCTGCGGTGGAAATGTTGATGGGAAAGATTTCAGAGGAGC 1350  
 Db 1258 GACCCACCCCACTGCGAAGGCTGCGGTGGAAATGTTGATGGGAAAGATTTCAGAGGAGC 1317  
 QY 1351 AAATTTAAAGTCTCCCTTGTGCGAAGAGACCTTCATGAAAGATGCGGGTGTGCTG 1410  
 Db 1318 AAATTTAAAGTCTCCCTTGTGCGAAGAGACCTTCATGAAAGATGCGGGTGTGCTG 1377  
 QY 1411 CCACCCCGTGAAGGCGAGAGGATGCCACCACTCCTGTGAGGTCCAGAGGCCCAAGGA 1470  
 Db 1378 CCACCCCGTGAAGGCGAGAGGATGCCACCACTCCTGTGAGGTCCAGAGGCCCAAGGA 1437  
 QY 1471 GGCTCTGGGGGAGCCCATGGGTGCGATGGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCT 1530  
 Db 1438 GGCTCTGGGGGAGCCCATGGGTGCGATGGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCT 1497  
 QY 1531 CCAAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGAGAGAGAGAAAGTCCAGACCGA 1590  
 Db 1498 CCAAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGAGAGAGAGAAAGTCCAGACCGA 1557  
 QY 1591 GCTGAGAGACTGGCATGTCTCCCATCCCGGTTGTGAGAAACAGAACTTCCGCTGAGAGCA 1650  
 Db 1558 GCTGAGAGACTGGCATGTCTCCCATCCCGGTTGTGAGAAACAGAACTTCCGCTGAGAGCA 1617  
 QY 1651 GAGTGCAACCAAGTGTAGGCGCCCAAGGCTGAAAGGCTCTCCCGGCAACCTTCCGCGCC 1710  
 Db 1618 GAGTGCAACCAAGTGTAGGCGCCCAAGGCTGAAAGGCTCTCCCGGCAACCTTCCGCGCC 1677  
 QY 1711 CCGGTTGTGATCGTGGCAGAGGTGGCTGTGTGGCATGCGGGAGAGAGAGGTGGCTC 1770  
 Db 1678 CCGGTTGTGATCGTGGCAGAGGTGGCTGTGTGGCATGCGGGAGAGAGAGGTGGCTC 1737  
 QY 1771 ATGATGATGTGTGTGTCCTCCGCTGAGATTTCAAGAGTGGCCGCTGTGAGACAGAGTGGC 1830  
 Db 1738 ATGATGATGTGTGTGTCCTCCGCTGAGATTTCAAGAGTGGCCGCTGTGTGAGACAGAGTGGC 1797  
 QY 1831 TTCCGTGTGTGGCGGGGATGAGACCAAGGTGGCTTGTGTGAGAGAGACAGAGGTGGCT 1890  
 Db 1798 TTCCGTGTGTGGCGGGGATGAGACCAAGGTGGCTTGTGTGAGAGAGACAGAGGTGGCT 1857  
 QY 1891 GGGGGGCGCCCTGAGACCTTTGATGAGAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1950  
 Db 1858 GGGGGGCGCCCTGAGACCTTTGATGAGAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917  
 QY 1951 CTTGAGAAATGATTAAGGAGACCGCTGAGAGCGCAGAGATCGGCGCTTACTAGATG 2010  
 Db 1918 CTTGAGAAATGATTAAGGAGACCGCTGAGAGCGCAGAGATCGGCGCTTACTAGATG 1977  
 QY 2011 CAGAGACCCCGCAGAG 2025  
 Db 1978 CAGAGACCCCGCAGAG 1992

RESULT 9

BC011048  
 LOCUS 2364 bp mRNA 11nec PRI 29 JUN-2004  
 DEFINITION Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 EMS, mRNA (CDNA clone MGC:16900 IMAGE:4340116), complete cds.  
 ACCESSION BC011048  
 VERSION BC011048.1 GI:15029674  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Joshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullan, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skliska, U., Smal, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)  
 12477932  
 2 (bases 1 to 2364)  
 REFERENCE  
 JOURNAL Strausberg, R.  
 AUTHORS Direct Submissions  
 TITLE Submitted (25-JUL-2001) National Institutes of Health, Mammalian  
 JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
 USA  
 REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Lou Staudt  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegue, H.,  
 Kowls, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAX Plate: 12 Row: P Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4885224.  
 Location/Qualifiers  
 1..2364  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:16900 IMAGE:4340116"  
 /issue\_type="Lymph, Lymphoma"  
 /clone\_lib="NIH MGC\_85"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 1..2364  
 /gene="EWSR1"

gene



Query Match	97.3% Score 1669.4; DB 12; Length 1971;
Best Local Similarity 99.9%; Pred. No. 0;	
Matches 1970; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY 37 ATGGCTTCACAGATTACAGTACCTATATGCCAAGCTGACGCGACGAGGGCTACAGTCT 96	
DB 1 ATGGCTTCACAGATTACAGTACCTATATGCCAAGCTGACGCGACGAGGGCTACAGTCT 60	
QY 97 TACACGCCCCCAAGCCCACTCAAGATATGCAACAACCCAGGGATATGGGCAACAAGC 156	
DB 61 TACACGCCCCCAAGCCCACTCAAGATATGCAACAACCCAGGGATATGGGCAACAAGC 120	
QY 157 TATGGAACTATGGACAGGCCCACTGAATGACGATATCCAGGGCTCAGACCACTGCAACC 216	
DB 121 TATGGAACTATGGACAGGCCCACTGAATGACGATATCCAGGGCTCAGACCACTGCAACC 180	
QY 217 TATGGAGAGACCGGCTATGCACTTCTTATGGAAGCGCTCCCACTGGTTATACTACTCA 276	
DB 181 TATGGAGAGACCGGCTATGCACTTCTTATGGAAGCGCTCCCACTGGTTATACTACTCA 240	
QY 277 ACTGCCCCCAAGGCAATACGCCAGCTGTCCAGGGGATATGGCACTGGTGCTTATGATACC 336	
DB 241 ACTGCCCCCAAGGCAATACGCCAGCTGTCCAGGGGATATGGCACTGGTGCTTATGATACC 300	
QY 337 ACCACTCTCTACAGTACCAACCAACGAGCCTCCTATGACGCTCAAGTGTGCAATATGGCACT 396	
DB 301 ACCACTCTCTACAGTACCAACCAACGAGCCTCCTATGACGCTCAAGTGTGCAATATGGCACT 360	
QY 397 CAGCGCTCTTATCCAGGCTATGGGACAGACCCAGACGCACTGCACTTACAAAGACCGCAG 456	
DB 361 CAGCGCTCTTATCCAGGCTATGGGACAGACCCAGACGCACTGCACTTACAAAGACCGCAG 420	
QY 457 GATGAAACAAAGCCCACTGAGACTAGTCAACTCTCATCTTACGACAGGGGGTTACAACAG 516	
DB 421 GATGAAACAAAGCCCACTGAGACTAGTCAACTCTCATCTTACGACAGGGGGTTACAACAG 480	
QY 517 CCCAGCCTAAGATATGGAAGAGTAATTAAGTTATCCCAAGTAACCTGGGAGGTAAACCC 576	
DB 481 CCCAGCCTAAGATATGGAAGAGTAATTAAGTTATCCCAAGTAACCTGGGAGGTAAACCC 540	
QY 577 ATGAGAGCAAGTCACTGACACCTCCATCCCTCCTACAGAGCTATTCCTTACACAGACCG 636	
DB 541 ATGAGAGCAAGTCACTGACACCTCCATCCCTCCTACAGAGCTATTCCTTACACAGACCG 600	
QY 637 ACTAGTTATGATCAGACAGATTACTCTCAGACAGAACCTATGGGCAACCGAGAGCTAT 696	
DB 601 ACTAGTTATGATCAGACAGATTACTCTCAGACAGAACCTATGGGCAACCGAGAGCTAT 660	
QY 697 GGAACAGAGTAGTACTATGTCACAAAGACGCTATGGGACAGAGCTCCCACTAAGTTAC 756	
DB 661 GGAACAGAGTAGTACTATGTCACAAAGACGCTATGGGACAGAGCTCCCACTAAGTTAC 720	
QY 757 CCACCCCAAACTGGATCTCTACAGCAAGCTCMAAGTCAATATAGCCAAACAGACGAGC 816	
DB 721 CCACCCCAAACTGGATCTCTACAGCAAGCTCMAAGTCAATATAGCCAAACAGACGAGC 780	
QY 817 TACGGGACAGAGATTCAATCCGACAGAACCCCACTAGTACATGGGTGTTATATGGGAG 876	
DB 781 TACGGGACAGAGATTCAATCCGACAGAACCCCACTAGTACATGGGTGTTATATGGGAG 840	

QY 877 GAGCTGAGGAGATTTTCCGGAACGAGAGAGAACCGAGCATGAGTGGCCCTGATTAACCGG 936  
 DB 841 GAGCTGAGGAGATTTTCCGGAACGAGAGAGAACCGAGCATGAGTGGCCCTGATTAACCGG 900  
 QY 937 GGCAGGGGAAAGGGGGGATTTGATCGTGAAGCATGAGCAGAGGTGGCGGGAGAGAGA 996  
 DB 901 GGCAGGGGAAAGGGGGGATTTGATCGTGAAGCATGAGCAGAGGTGGCGGGAGAGAGA 960  
 QY 997 CGGGGTGGAATGGCAGCGCTGAGAGCAGAGTGGCTTCAATTAAGCTTGATGACCAATG 1056  
 DB 961 CGGGGTGGAATGGCAGCGCTGAGAGCAGAGTGGCTTCAATTAAGCTTGATGACCAATG 1020  
 QY 1057 GATGAAGACCAAGATCTTGAATCTAGGCGCACTGTAGATCAGATGAAGACTGACAC 1116  
 DB 1021 GATGAAGACCAAGATCTTGAATCTAGGCGCACTGTAGATCAGATGAAGACTGACAC 1080  
 QY 1117 AGTGCATTTATGTACAAAGATTAATATGACAGTGTGACTAGATGATCTGACAGCTTC 1176  
 DB 1081 AGTGCATTTATGTACAAAGATTAATATGACAGTGTGACTAGATGATCTGACAGCTTC 1140  
 QY 1177 TTTAAGAGTGTGGGGTGTGTTAGATGAACAAGAACTGGGCAACCCATGATCAATC 1236  
 DB 1141 TTTAAGAGTGTGGGGTGTGTTAGATGAACAAGAACTGGGCAACCCATGATCAATC 1200  
 QY 1237 TACCTGGCAAGGAAACAGGAAAGCCAAAGCGCATGCCACATGCTCTATGAAGACCA 1296  
 DB 1201 TACCTGGCAAGGAAACAGGAAAGCCAAAGCGCATGCCACATGCTCTATGAAGACCA 1260  
 QY 1297 CCCACTGCCAAGCGCTGCGTGAATGTGTTAGTGGAAAGATTTTCAAGGAGCAACTT 1356  
 DB 1261 CCCACTGCCAAGCGCTGCGTGAATGTGTTAGTGGAAAGATTTTCAAGGAGCAACTT 1320  
 QY 1357 AAGAGTCTCCCTGTGCGAAGAAAGCTTCAATGAACGATGCGGGTGTGCTGCCACC 1416  
 DB 1321 AAGAGTCTCCCTGTGCGAAGAAAGCTTCAATGAACGATGCGGGTGTGCTGCCACC 1380  
 QY 1417 CGTGAAGGCGAGGCAATGCCAACCACTCCGTGAGATGTCAGAGAGCCCAAGAGGTCTT 1476  
 DB 1381 CGTGAAGGCGAGGCAATGCCAACCACTCCGTGAGATGTCAGAGAGCCCAAGAGGTCTT 1440  
 QY 1477 GGGGAGCCCATGGGTGCGATGAGAGCGGTGAGAGAGATAGAGAGGCTTCCCTCCAGA 1536  
 DB 1441 GGGGAGCCCATGGGTGCGATGAGAGCGGTGAGAGAGATAGAGAGGCTTCCCTCCAGA 1500  
 QY 1537 GGAAGCCCGGGGTTCCGAGGAGAACCCCTCTGAGAGAGAAACGTCCAGCAACGAGCTGA 1596  
 DB 1501 GGAAGCCCGGGGTTCCGAGGAGAACCCCTCTGAGAGAGAAACGTCCAGCAACGAGCTGA 1560  
 QY 1597 GACTGGCAGTGTCCCAATCCGGGGTGTGGAACAAGAACTTCCGCTGAGAGACAGATGC 1656  
 DB 1561 GACTGGCAGTGTCCCAATCCGGGGTGTGGAACAAGAACTTCCGCTGAGAGACAGATGC 1620  
 QY 1657 AACCAAGTAAAGGCCCCAAGGCTGAAGGCTTCTCCGCGCAACCTTCCGCCCCGGGT 1716  
 DB 1621 AACCAAGTAAAGGCCCCAAGGCTGAAGGCTTCTCCGCGCAACCTTCCGCCCCGGGT 1680  
 QY 1717 GGTGATCTGTGAGCAAGGTGAGCCCTGTGAGCAATGCGGGAGAGAAAGGTGAGCTCATGAT 1776  
 DB 1681 GGTGATCTGTGAGCAAGGTGAGCCCTGTGAGCAATGCGGGAGAGAAAGGTGAGCTCATGAT 1740  
 QY 1777 CGTGTGTGTCCCGGTGGAATGTTCAAGAGTGTGCGGTGTGAGACAGAGGTGCTTCCT 1836  
 DB 1741 CGTGTGTGTCCCGGTGGAATGTTCAAGAGTGTGCGGTGTGAGACAGAGGTGCTTCCT 1800  
 QY 1837 GGTGAGCCGGGCGATGAGCCGAGGAGGCTTGTGTGAGAGAAAGAGAGGTGAGCCCTGGGGG 1896  
 DB 1801 GGTGAGCCGGGCGATGAGCCGAGGAGGCTTGTGTGAGAGAAAGAGAGGTGAGCCCTGGGGG 1860  
 QY 1897 CCCCTGAGACTTTGATGAGAAACAGATGAGAGAGAGAAAGAGAGAGCTGAGAGACCTGGA 1956  
 DB 1861 CCCCTGAGACTTTGATGAGAAACAGATGAGAGAGAGAAAGAGAGAGCTGAGAGACCTGGA 1920  
 QY 1957 AAAATGATTAAGGCGAGCACCGTCAGAGAGCGCAGAGATGGCCCTTACTAG 2007

Db

1921 AAAATGATTAAGGCGAGCACCGTCAGAGCGCAGAGATGGCCCTTACTTG 1971

## RESULT 11

AK056309

LOCUS

## DEFINITION

AK056309 2189 bp mRNA linear PRI 30-JAN-2004  
 Homo sapiens cDNA FLJ1747 f1s, clone NT2R12007377, highly similar  
 to RNA-BINDING PROTEIN EWS.

## ACCESSION

AK056309.1 GI:16551673

## VERSION

AK056309.1 GI:16551673

## KEYWORDS

oligo capping; f1s (full insert sequence).

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Oca, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Makatsugu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, F., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuna, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, E., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, M., Omura, Y., Abe, K., Kamihara, K., Katsura, N., Sato, K., Tanikawa, M., Yamazaki, M., Nimomiyu, K., Ishida, T., Yamashita, H., Hirakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoda, M., Houta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mueselino, K., Yuki, H., Ohnima, A., Sasaki, N., Aotsuka, S., Yoshioka, Y., Matsunawa, H., Ichihara, T., Shionata, N., Sano, S., Moriya, S., Momiya, H., Sato, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kamegaki, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Taniguchi, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Omori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunaga, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togsashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, T., Nagase, T., Nomura, N., Kikuchi, H., Masuko, Y., Yamashita, K., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S.

## TITLE

Complete sequencing and characterization of 21,243 full-length human cDNAs

## JOURNAL

Nat. Genet. 36 (1), 40-45 (2004)

## PUBMED

14702039

## REFERENCE

Ishibashi, T., Kanehori, K., Yoda, M., Watanabe, S., Ishida, S., Ono, Y., Houta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makatsugu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Matsuno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuna, M., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahari, K., Masuko, Y., Nagai, K., and Isogai, T.

## AUTHORS

NEBO human cDNA sequencing project

## JOURNAL

Unpublished

## REFERENCE

3 (bases 1 to 2189)

## AUTHORS

Isogai, T., Otsuki, T., and Sugiyama, T.

## TITLE

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 222-0812, Japan (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

## JOURNAL

NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- &amp; 3'-end one pass sequencing: RAB,

## COMMENT





QY 1618 GGTGTGGAACAGAACTTCGCTGAGAAAGAGTGAACAGTGTAGAGCCCAAG 1677  
DB 1618 GGTGTGGAACAGAACTTCGCTGAGAAAGAGTGAACAGTGTAGAGCCCAAG 1677  
QY 1678 CTTGAAGGCTTCTCCGCCCACTTTCCGCCCGGGGTGTATCTGTGAGAGTGGC 1737  
DB 1678 CTTGAAGGCTTCTCCGCCCACTTTCCGCCCGGGGTGTATCTGTGAGAGTGGC 1737  
QY 1738 CTTGTGTGAGTGGGGGAGAGAGTGGCTCATGTATCTGTGTGTCCCGGTGAGT 1797  
DB 1738 CTTGTGTGAGTGGGGGAGAGAGTGGCTCATGTATCTGTGTGTCCCGGTGAGT 1797  
QY 1798 TTCAAGAGTGGCCGTGTGTGAGAGAGAGTGTCTTCCGTGTGGCCGGGATGAGC 1857  
DB 1798 TTCAAGAGTGGCCGTGTGTGAGAGAGAGTGTCTTCCGTGTGGCCGGGATGAGC 1857  
QY 1858 GGTGGCTTTGGTGTGAGAGAGAGAGTGGCCCTTGGGGGGCCCTGTGACCTTTGATGAA 1917  
DB 1858 GGTGGCTTTGGTGTGAGAGAGAGAGTGGCCCTTGGGGGGCCCTGTGACCTTTGATGAA 1917  
QY 1918 CAGATGGAGAGAGAGAGAGAGAGAGAGTGGAGAGCTTGGAGAGAGTGAAGAGAGAGC 1977  
DB 1918 CAGATGGAGAGAGAGAGAGAGAGAGAGTGGAGAGCTTGGAGAGAGTGAAGAGAGAGC 1977  
QY 1978 CGTCAGAGAGAGAGAGAGTGGCCCTTACTAGATGAGAGAGAGAGAGAGAGAGAG 2025  
DB 1978 CGTCAGAGAGAGAGAGAGTGGCCCTTACTAGATGAGAGAGAGAGAGAGAGAGAG 2025

RESULT 12  
LOCUS CO730352 2326 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 16286 from Patent WO02068579.  
VERSION CO730352 GI:42303963  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Rites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
theroeof

JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;  
PE Corporation (NY) (US)

FEATURES  
source  
1. .2326  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 97.0%; Score 1964.4; DB 6; Length 2326;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1976; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 49 GATTACAGTACTATAGCAAGCTGACGCGCAGAGGGCTACAGTGTCTTACCGCCAG 108  
DB 8 GATTACAGTACTATAGCAAGCTGACGCGCAGAGGGCTACAGTGTCTTACCGCCAG 67  
QY 109 CCCACTCAGAGATATGACAGACCAACGAGCATATGGGCAACAAAGCTATGAACTAT 168  
DB 68 CCCACTCAGAGATATGACAGACCAACGAGCATATGGGCAACAAAGCTATGAACTAT 127  
QY 169 GAGCAGGCGCATGTGCTAGTACCTATACCCAGGCTCAGACCTGAACTCTATGGGAGACC 228  
DB 128 GAGCAGGCGCATGTGCTAGTACCTATACCCAGGCTCAGACCTGAACTCTATGGGAGACC 187  
QY 229 GCTGTATGCACTTTATATGAGACAGCTCCCACTGTGTATATCTCCAACTGCCCCAG 288  
DB 188 GCTGTATGCACTTTATATGAGACAGCTCCCACTGTGTATATCTCCAACTGCCCCAG 247

QY 289 GCATACAGCCAGCCTGTCCAGGGGTATGAGACTGTGCTTATGATACCAACACTGTACA 348  
DB 248 GCATACAGCCAGCCTGTCCAGGGGTATGAGACTGTGCTTATGATACCAACACTGTACA 307  
QY 349 GTCAACACCAACCCAGGCTCTTATGACAGTCACTGATATGAGCACTGAGCTGTATAT 408  
DB 308 GTCAACACCAACCCAGGCTCTTATGACAGTCACTGATATGAGCACTGAGCTGTATAT 367  
QY 409 CCAGCTTATGGGAG 468  
DB 368 CCAGCTTATGGGAG 427  
QY 469 CCCACTCAGAGATATGACAGTCACTGATATGAGCACTGAGCACTGAGCTGTATAT 528  
DB 428 CCCACTCAGAGATATGACAGTCACTGATATGAGCACTGAGCACTGAGCTGTATAT 487  
QY 529 TATGAGCAGAGTAACTACATGTTATCCCAAGGTACTGGAGCTTACCCCATGAGAGAGT 588  
DB 488 TATGAGCAGAGTAACTACATGTTATCCCAAGGTACTGGAGCTTACCCCATGAGAGAGT 547  
QY 589 ACTGACCTTCATCTTACCTCTTACCAAGCTATTTCTTACAGAGCCGACTAGTTATGAT 648  
DB 548 ACTGACCTTCATCTTACCTCTTACCAAGCTATTTCTTACAGAGCCGACTAGTTATGAT 607  
QY 649 CAGAGCAGTAACTCTGAG 708  
DB 608 CAGAGCAGTAACTCTGAG 667  
QY 709 AGCTATGATCAACAAAG 768  
DB 668 AGCTATGATCAACAAAG 727  
QY 769 GATCTCTACAG 828  
DB 728 GATCTCTACAG 787  
QY 829 AGTTCACTTCCGAG 888  
DB 788 AGTTCACTTCCGAG 847  
QY 889 TTTTCCGAG 948  
DB 848 TTTTCCGAG 907  
QY 949 GGGGAGATTGATGATGAG 1008  
DB 908 GGGGAGATTGATGATGAG 967  
QY 1009 GGGGAGATTGATGATGAG 1068  
DB 968 GGGGAGATTGATGATGAG 1027  
QY 1069 GATCTGATCTAG 1128  
DB 1028 GATCTGATCTAG 1087  
QY 1129 GTACAGAGATTAAATAGCAGTGTGATCTGATGATCTGAGAGAGTCTTTAAGAGTGT 1188  
DB 1088 GTACAGAGATTAAATAGCAGTGTGATCTGATGATCTGAGAGAGTCTTTAAGAGTGT 1147  
QY 1189 GGGGTTTAAAGTAAAG 1248  
DB 1148 GGGGTTTAAAGTAAAG 1207  
QY 1249 GAAACAGAGAAAGCCAAAGGCGATGCGACAGTGTCTTATGAAAGAGAGAGAGAGAG 1308  
DB 1208 GAAACAGAGAAAGCCAAAGGCGATGCGACAGTGTCTTATGAAAGAGAGAGAGAGAG 1267  
QY 1309 GCTGCGGTGAAAGTGTGTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368  
DB 1268 GCTGCGGTGAAAGTGTGTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1327



QY 1369 GCTCGAAGAAAGCTCCAAATGAAACAGNATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGA 1428  
 DB 1328 GCTCGAAGAAAGCTCCAAATGAAACAGATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGA 1387  
 QY 1429 GGCATGCGACACCACTCCGTGAGGTCCAGAGAGCCAGAGAGTCTCTGGGGAGACCATG 1488  
 DB 1388 GGCATGCGACACCACTCCGTGAGGTCCAGAGAGCCAGAGAGTCTCTGGGGAGACCATG 1447  
 QY 1489 GGTGCGATGAGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAGAGAGACCCCGGGGT 1548  
 DB 1448 GGTGCGATGAGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAGAGAGACCCCGGGGT 1507  
 QY 1549 TCCCGAGAGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGACCTGGCAAGT 1608  
 DB 1508 TCCCGAGAGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGACCTGGCAAGT 1567  
 QY 1609 CCCAATCCGGATTGTGAGAAACAGAACTTGCCTGAGAAACAGAGTCAACCAAGTAAAG 1668  
 DB 1568 CCCAATCCGGATTGTGAGAAACAGAACTTGCCTGAGAAACAGAGTCAACCAAGTAAAG 1627  
 QY 1669 GCCCAAAAGCTTGAAGGCTTCTCCGCGCACCTTTCCGCCCCCGGGTGTGATCGTGC 1728  
 DB 1628 GCCCAAAAGCTTGAAGGCTTCTCCGCGCACCTTTCCGCCCCCGGGTGTGATCGTGC 1687  
 QY 1729 AGAGTGGCCCTGTGTGATGAGAGAGGAGAGAGAGTGGCTCAGATGATGATGATGATG 1788  
 DB 1688 AGAGTGGCCCTGTGTGATGAGAGAGGAGAGAGAGTGGCTCAGATGATGATGATGATG 1747  
 QY 1789 GGTGGAATGTTCAAGAGTGGCGCTGTGTGAGACAGAGAGTGGCTT-CCGTGATGAGCGGG 1847  
 DB 1748 GGTGGAATGTTCAAGAGTGGCGCTGTGTGAGACAGAGAGTGGCTT-CCGTGATGAGCGGG 1807  
 QY 1848 CATGAGACCGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGAC 1907  
 DB 1808 CATGAGACCGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGAC 1867  
 QY 1908 TTTGATGGAACAGATGAG 1967  
 DB 1868 TTTGATGGAACAGATGAG 1927  
 QY 1968 AGGCGAGACCGCTGAG 2025  
 DB 1928 AGGCGAGACCGCTGAG 1985  
  
 RESULT 13  
 AC137500/c 193559 bp DNA linear HTG 20-NOV-2002  
 LOCUS Homo sapiens chromosome 16 clone RP11-343G17, WORKING DRAFT  
 DEFINITION  
 SEQUENCE, 2 unordered pieces.  
 AC137500.1 GI:25139894  
 ACCESSION  
 VERSION HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 193559)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Sequencing of Human Chromosome 16  
 REFERENCE 2 (bases 1 to 193559)  
 TITLE DOE Joint Genome Institute.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (20-NOV-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 538238

Center clone name: RP11-11\_343G17  
 -----  
 Summary Statistics  
 Consensus quality: 192718 bases at least Q40  
 Consensus quality: 192915 bases at least Q30  
 Consensus quality: 193071 bases at least Q20  
 Estimated insert size: 160000; agarose-fp estimation  
 Estimated insert size: 193459; sum-of-contigs estimation  
 Quality coverage: 12.34 in Q20 bases; agarose-fp estimation  
 Quality coverage: 10.21 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 96955: contig of 96955 bp in length  
 \* 96956 97055: gap of unknown length  
 \* 97056 193559: contig of 96504 bp in length.  
 Location/Qualifiers  
 source  
 1. 193559  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-343G17"  
 /clone\_lib="RP11 human BAC library 11"  
  
 ORIGIN  
 Query Match 93.1%; Score 1884.4; DB 2; Length 193559;  
 Best Local Similarity 96.4%; Pred. No. 0;  
 Matches 1949; Conservative 0; Mismatches 71; Indels 2; Gaps 2;  
  
 QY 4 GGAAGTTGAG 63  
 DB 31039 GGAAGTTGAG 30980  
 QY 64 AGCCAAAGCTGACGCGAT 123  
 DB 30979 AGCCAAAGCTGACGCGAT 30920  
 QY 124 GCAAGAGACCAAGAGAGATATGAT 183  
 DB 30919 GCAAGAGACCAAGAGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 30860  
 QY 184 GTCAGCTATACCAAGAT 243  
 DB 30859 GTCAGCTATACCAAGAT 30800  
 QY 244 TATGAGACAGCTCCCACTGCTTATATCTATCTCAACTGCCCCCAAGAGATACAGCCAGCT 303  
 DB 30799 TATGAGACAGCTCCCACTGCTTATATCTATCTCAACTGCCCCCAAGAGATACAGCCAGCT 30740  
 QY 304 GTCAGAGGGATATGAT 363  
 DB 30739 GTCAGAGGGATATGAT 30680  
 QY 364 GCTCTCTATGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423  
 DB 30679 GCTCTCTATGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 30620  
 QY 424 CAGCCAGACCACTGACCTTACAGAT 483  
 DB 30619 CAGCCAGACCACTGACCTTACAGAT 30560  
 QY 484 CAACCTCAATCTGACAGAT 543  
 DB 30559 CAACCTCAATCTGACAGAT 30500  
 QY 544 TACAGTTATCCAGAGTACTGAGAGAGATCCCAATGAGAGAGAGAGAGAGAGAGAGAGAGAT 603

Db 30499 TGCAGTTATCCCGAGGTACCTGGGAGCTACTGCAATGACGACGATCACCGGACCTCCATCC 30440  
 QY 604 TACCTCTCTACGAGCTATTCTCTTACACAGCCGACCTAGTTATGATCAAGAGCTTACTCT 663  
 Db 30439 TACCTCTCTACGAGCTATTCTCTTACACAGCCGACCTAGTTATGATCAAGAGCTTACTCT 30380  
 QY 664 CACAGAAACACTATGGGCAACCGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 723  
 Db 30379 CACAGAAACACTATGGGCAACCGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 30320  
 QY 724 AGCAGCTATGGGCAACCGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 783  
 Db 30319 AGCAGCTATGGGCAACCGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 30260  
 QY 784 GCTCCAACTATGAGTCAACAGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 843  
 Db 30259 GCTCCAACTATGAGTCAACAGAGAGCTATGGACAGAGAGTACTATGGTCAACAA 30200  
 QY 844 GACCAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 903  
 Db 30199 GACCAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 30140  
 QY 904 GAGAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 963  
 Db 30139 GAGAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 30080  
 QY 964 GAGAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 1023  
 Db 30079 GAGAACCCGAGTATGGGCTTATGGGCAAGGCTGAGAGATTTTCCGACCGACGA 30020  
 QY 1024 CGAGTGGCTTCAATAAGCTGATGAGCCATGAGTGAAGACCAATCTTATGATCTTATGAGC 1083  
 Db 30019 CGAGTGGCTTCAATAAGCTGATGAGCCATGAGTGAAGACCAATCTTATGATCTTATGAGC 29960  
 QY 1084 CCACTGTAGATCAGATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 1143  
 Db 29959 CCACTGTAGATCAGATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 29900  
 QY 1144 GACAGTGTGATCTTATGATCTTATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 1203  
 Db 29899 GACAGTGTGATCTTATGATCTTATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 29840  
 QY 1204 AACCAAGAACTGGGCAACCATGATCACAATCTACTGACAAAGAAACAGAAAGCC 1263  
 Db 29839 AACCAAGAACTGGGCAACCATGATCACAATCTACTGACAAAGAAACAGAAAGCC 29780  
 QY 1264 AAAGCGATGCAAGTGTCTTATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 1323  
 Db 29779 AAAGCGATGCAAGTGTCTTATGAGACCTGACACAGTGCATTTATGATCAAGATTAAT 29720  
 QY 1324 TTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTTGTCTGGAAGAGCCT 1383  
 Db 29719 TTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTTGTCTGGAAGAGCCT 29660  
 QY 1384 CCAATGAACGATGAGGAGTGTCTGCAACCCCGTGAAGGAGAGAGATCCACCAACCA 1443  
 Db 29659 CCAATGAACGATGAGGAGTGTCTGCAACCCCGTGAAGGAGAGAGATCCACCAACCA 29600  
 QY 1444 CTCCTGAGAGTCCAGAGGCTCTGAGGAGACCTATGGGCTGATGGAGG 1503  
 Db 29599 CTCCTGAGAGTCCAGAGGCTCTGAGGAGACCTATGGGCTGATGGAGG 29540  
 QY 1504 CTTGAGAGAGTCCAGAGGCTCTTCCCAAGAGACCCCGGAGTCTCCGAGGAAACCC 1563  
 Db 29539 CTTGAGAGAGTCCAGAGGCTCTTCCCAAGAGACCCCGGAGTCTCCGAGGAAACCC 29480  
 QY 1564 TCTGAGAGAGTCCAGAGGCTCTTCCCAAGAGACCCCGGAGTCTCCGAGGAAACCC 1623  
 Db 29479 TCTGAGAGAGTCCAGAGGCTCTTCCCAAGAGACCCCGGAGTCTCCGAGGAAACCC 29420  
 QY 1624 GGAAGACAGAACTTCCCTGAGAGACAGAGTGCACCAAGTGTAAAGCCCAAGCTGAA 1683  
 Db 29419 GGAAGACAGAACTTCCCTGAGAGACAGAGTGCACCAAGTGTAAAGCTGAA 29360

QY 1684 GGCCTTCTCCCGCACCCTTTCCGCCCGCGGTGATGTCGACAGAGGTGGCCTGCT 1743  
 Db 29359 GGCCTTCTCCCGCACCCTTTCCGCCCGCGGTGATGTCGACAGAGGTGGCCTGCT 29300  
 QY 1744 GGCATCGGGGAGAGAGAGGTGGCTCATGAGTGTGATGTCCTCGGTGAAATTTTCA 1803  
 Db 29299 GGCATCGGGGAGAGAGAGGTGGCTCATGAGTGTGATGTCCTCGGTGAAATTTTCA 29240  
 QY 1804 GGTGGCCGTGTTGAGACAGAGGTGGCTTCCGTGTGTCGCGGAGATGACCCGAGTGGC 1863  
 Db 29239 GGTGGCCGTGTTGAGACAGAGGTGGCTTCCGTGTGTCGCGGAGATGACCCGAGTGGC 29181  
 QY 1864 TTTGTTGAGAGAGAGAGAGGTGGCTTCCGTGTGTCGCGGAGATGACCCGAGTGGC 1923  
 Db 29180 TTTGTTGAGAGAGAGAGAGGTGGCTTCCGTGTGTCGCGGAGATGACCCGAGTGGC 29122  
 QY 1924 GAGAGAAAGAGAGAGAGAGTGGACCTGGAAGAAATGATTAAGCGGACCCCTCAG 1983  
 Db 29121 GAGAGAAAGAGAGAGAGAGTGGACCTGGAAGAAATGATTAAGCGGACCCCTCAG 29062  
 QY 1984 GAGCGAGAGATGCGCCTTACTAGATGACAGAGACCCGCGAGA 2025  
 Db 29061 GAGCGAGAGATGCGCCTTACTAGATGACAGAGACCCGCGAGA 29020

RESULT 14  
 AC018774  
 LOCUS  
 DEFINITION  
 Homo sapiens clone RP11-2093, WORKING DRAFT SEQUENCE, 3 unordered  
 pieces.  
 AC018774 GI:7107978  
 VERSION  
 HTG; HTGS PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 155815)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Castele,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearrellano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehocsky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
 Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,  
 Meldrum,J., Menus,L., Morrow,J., Naylor,J., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N.,  
 Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,  
 Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
 Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.  
 and Zody,M.  
 Direct Submision  
 Submitted (19-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 28, 2000 this sequence version replaced gi:692312.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/kw/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence submissions@genome.wi.mit.edu  
 ----- Project Information





QY 424 CAGCCAGCAGCACTGCACTTCAAGACCGGAGATGAAAACAAGCCCACTGAGACTAGT 483  
 DB 157205 CAGCCATGACCTGCACTTCAAGACCTGAGATGAAAACAAGCCCACTGAGACTAGT 157146  
 QY 484 CAACCTCAATCTAGCACAAGGGGGTTTACCAACAGCCAGCCTAGATATGACAGAGTAC 543  
 DB 157145 CAACCTCAATCTAGCACAAGGGGGTTTACCAACAGCCAGCCTAGATATGACAGAGTAC 157086  
 QY 544 TACAGTATATCCCAAGGATCTGAGGAGCTACCCATGACAGCCAGTCACTGCACTCCATCC 603  
 DB 157085 TGAAGTATATCCCAAGGATCTGAGGAGCTACCCATGACAGCCAGTCACTCCATCC 157026  
 QY 604 TACCTCTCTACAGACTATTCCTCTACACAGCCAGCTAGTATGATGACAGAGTACTCT 663  
 DB 157025 TACCTCTCTACAGACTATTCCTCTACACAGCCAGCTAGTATGATGACAGAGTACTCT 156966  
 QY 664 CAGCAGAACACTATGAGGCAACCGAGCAGCTATGACAGCAGAGTACTATGCTCAACAA 723  
 DB 156965 CAGCAGAACACTATGAGGCAACCGAGCAGCTATGACAGCAGAGTACTATGCTCAACAA 156906  
 QY 724 AGCAGCTATGAGGAGCAGCCTCCCACTAGTTACCCAACTGATCTCTACAGCCAA 783  
 DB 156905 AGCAGCTATGAGGAGCAGCCTCCCACTAGTTACCCAACTGATCTCTACAGCCAA 156846  
 QY 784 GCTCAAGTCAATATAGCAACAGAGAGCAGCTACGAGGAGCAGAGTTCCTCCAGCAG 843  
 DB 156845 GCTCAAGTCAATATAGCAACAGAGAGCAGCTACGAGGAGCAGAGTTCCTCCAGCAG 156786  
 QY 844 GACCACCCCAAGTACATGAGGTTTATGAGGAGAGTCTGAGAGATTTTCCGACCAAGA 903  
 DB 156785 GACCACCCCAAGTACATGAGGTTTATGAGGAGAGTCTGAGAGATTTTCCGACCTAGGA 156726  
 QY 904 GAGAACCCGAGCAGTATGAGGCTCTGATTAACCGGGGAGGGGGAAGGGGGATTTGATCGT 963  
 DB 156725 GAGAACCCGAGCAGTATGAGGCTCTGATTAACCGGGGAGGGGGAAGGGGGATTTGATCGT 156666  
 QY 964 GAGAGCAGTACAGAGGATGAGGCGGAGAGAGAGACGCGGTGAAATGCGCAGCGCTGAGAG 1023  
 DB 156665 GAGAGCAGTACAGAGGATGAGGCGGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGG 156606  
 QY 1024 CGAGGTGCTTCAATTAAGCTGTGTGAACCATGAGTAAAGAACCATCTTGAATCTAGGC 1083  
 DB 156605 CGAGGTGCTTCAATTAAGCTGTGTGAACCATGAGTAAAGAACCATCTTGAATCTAGGC 156546  
 QY 1084 CCACCTGTATATCCAGTATGAAGACTCTGACAAAGTGCATTTATGTAAGAGATTTAAAT 1143  
 DB 156545 CCACCTGTATATCCAGTATGAAGACTCTGACAAAGTGCATTTATGTAAGAGATTTAAAT 156486  
 QY 1144 GACAGTGTGACTAGATGATCTGCGAGACTTCTTTAAGCAGTGTGAGGTTGTTAAGATG 1203  
 DB 156485 GACAGTGTGACTAGATGATCTGCGAGACTTCTTTAAGCAGTGTGAGGTTGTTAAGATG 156426  
 QY 1204 AACAGAGAACTGAGGCAACCATGATCCATCTTACCTGAGCAAGAAACAGAAAGCC 1263  
 DB 156425 AACAGAGAACTGAGGCAACCATGATCCATCTTACCTGAGCAAGAAACAGAAAGCC 156366  
 QY 1264 AAAGGCATGCGACAGTGTCTTATGAAGACCAACCCACTGCAAGGCTGCGGTGAAATGG 1323  
 DB 156365 AAAGGCATGCGACAGTGTCTTATGAAGACCAACCCACTGCAAGGCTGCGGTGAAATGG 156306  
 QY 1324 TTTGATGGAAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGCGAAGAAAGCCT 1383  
 DB 156305 TTTGATGGAAGAAATTTTCAAGGAGCAAACTTAAAGTCTCTTGTGCGAAGAAAGCCT 156246  
 QY 1384 CCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGCAGAGCATGCCACCA 1443  
 DB 156245 CCAATGAACAGTATGCGGGGTGTATGCAACCCCGTGAAGGAGCAGAGCATGCCACCA 156186  
 QY 1444 CTCCGTGAGAGGTCCAGAGAGCCAGAGAGTCTTGGGGGAGCCCATGGGTCCGATGGAGGC 1503  
 DB 156185 CTCCGTGAGAGGTCCAGAGAGCCAGAGAGTCTTGGGGGAGCCCATGGGTCCGATGGAGGC 156126

QY 1504 CGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCGAGGAAACCC 1563  
 DB 156125 CGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGTTCGAGGAAACCC 156066  
 QY 1564 TCTGAGAGAGAAACGTTCACACACGAGCTGAGACTGCACTGCTCCAAATCCGGGTTGT 1623  
 DB 156065 TCTGAGAGAGAAACGTTCACACACGAGCTGAGACTGCACTGCTCCAAATCCGGGTTGT 156006  
 QY 1624 GGAACCAAGAACTTGGCTGAGAGACAGAGTGAACAGATGAAGCCCAAGAGCTGAA 1683  
 DB 156005 GGAACCAAGAACTTGGCTGAGAGACAGAGTGAACAGATGAAGCTCAAGAGCTGAA 155946  
 QY 1684 GGCCTCTCCGCACTTCTCCGCCCCGGGTGTGATCTGTGAGAGAGTGGCCCTGTGT 1743  
 DB 155945 GGCCTCTCCGCACTTCTCCGCCCCGGGTGTGATCTGTGAGAGAGTGGCCCTGTGT 155886  
 QY 1744 GGCATGCGGAGAGAAAGAGTGGCTCATGATCTGTGTGTCTCCGAGTATGTTTCA 1803  
 DB 155885 GGCATGCGGAGAGAAAGAGTGGCTCATGATCTGTGTGTCTCCGAGTATGTTTCA 155826  
 QY 1804 GGTGGCCGTGTGAGAGACAGAGTGGCTTCCGTGTGTGTCGCGGGCANTGACCGAGGTGGC 1863  
 DB 155825 GGTGGCCGTGTGAGAGACAGAGTGGCTTCCGTGTGTGTCGCGGGCANTGACCGAGGTGGC 155767  
 QY 1864 TTTGTGAGAGAAAGAGAGTGGCTTCCGAGGAGCCCTGAGACTTGTGATGAAACAGATG 1923  
 DB 155766 TTTGTGAGAGAAAGAGAGTGGCTTCCGAGGAGCCCTGAGACTTGTGATGAAACAGATG 155708  
 QY 1924 GAGAGAAAGAGAGAGAGAGTGGAGAGCTTGGAAAAATGATTAAGCCGAGCAGCTGAG 1983  
 DB 155707 GAGAGAAAGAGAGAGAGAGTGGAGAGAGCTTGGAAAAATGATTAAGCCGAGCAGCTGAG 155648  
 QY 1984 GAGCGAGAGATGGGCTTACTAGATGACAGAGACCCCGGAGA 2025  
 DB 155647 GAGCGAGAGATGGGCTTACTAGATGACAGAGACCCCGGAGA 155606

Search completed: February 20, 2005, 11:50:27  
 Job time : 8507.87 secs



a process for producing a polypeptide; (4) an isolated polypeptide comprising: (a) an amino acid sequence encoded by any of the above nucleotide sequences; (b) an amino acid sequence encoded by the full-length coding region of the above nucleotide sequences; or (c) a sequence having at least 80% identical to (a) or (b); (5) a chimeric polypeptide comprising the above polypeptide fused to a heterologous polypeptide; (6) an isolated antibody that binds to the above polypeptide; (7) a process for producing the antibody; (8) an isolated oligopeptide that binds to the above polypeptide; (9) a tumour-associated antigenic target (TAT) binding organic molecule that binds to the above polypeptide; (10) a composition of matter comprising the above (chimeric) polypeptide, antibody, oligopeptide or TAT binding organic molecule, in combination with a carrier; (11) an article of manufacture comprising a container and the composition of matter contained within the container; (12) methods of inhibiting the growth of a cell that expresses the above protein, where the growth of the cell is at least in part dependent upon a growth potentiating effect of the above protein; (13) a method of therapeutically treating a mammal having a cancerous tumour comprising cells that express the above protein; (14) a method of determining the presence of a protein in a sample suspected of containing the protein described above; (15) methods of diagnosing the presence of a tumour in a mammal; (16) a method for treating or preventing a cell proliferative disorder associated with increased expression or activity of the above protein; and (17) a method of binding an antibody, oligopeptide or organic molecule to a cell that expresses the protein described above. The TAT sequences have cytotoxic activities, and can be used in gene therapy. The composition and methods are useful for diagnosing, preventing or treating cancer. The composition is also used for preparing a medicament for the therapeutic treatment or diagnostic detection of a cell proliferative disorder or cancer. The present sequence represents a human TAT cDNA sequence from the present invention.

**SQ** Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match	100.0%;	Score 2025;	DB 13;	Length 2372;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2025; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	GAACGACGTTGAGAGAAACAGAGAGAAAGAGAAAATGGCGTTCACAGATTCAGATACC	60
Db	8	GAACGACGTTGAGAGAACAGAGAGAAAGAGAAAATGGCGTTCACAGATTCAGATACC	67
QY	61	TATTAACCAAGCTGCAGCGCGACAGAGGGCTACAGTGCCTTACACCGGCCACCTCAAGGA	120
Db	68	TATTAACCAAGCTGCAGCGCGACAGAGGGCTACAGTGCCTTACACCGGCCACCTCAAGGA	127
QY	121	TATGCACAGACCAACCCAGGCACTATGGGCAACAAGCTATGAACTTATGAGACGCCACT	180
Db	128	TATGCACAGACCAACCCAGGCACTATGGGCAACAAGCTATGAACTTATGAGACGCCACT	187
QY	181	GATGTACGCTATTAACCAAGGCTCAGACCACTGCACCTATGGGCAAGCCGCTTATGCAACT	240
Db	188	GATGTACGCTATTAACCAAGGCTCAGACCACTGCACCTATGGGCAAGCCGCTTATGCAACT	247
QY	241	TCCTATGACAGGCGTCCCACTGGTTTATCTACTCCAACTGCCCCCAGGCACTACAGCCAG	300
Db	248	TCCTATGACAGGCGTCCCACTGGTTTATCTACTCCAACTGCCCCCAGGCACTACAGCCAG	307
QY	301	CCTGTTCAGGGGGTATGGCACTGGTGTCTTATGATACCAACACTGCTACAGTACCAACCAAC	366
Db	308	CCTGTTCAGGGGGTATGGCACTGGTGTCTTATGATACCAACACTGCTACAGTACCAACCAAC	367
QY	361	CAGGCGCTCTTATGCACTAGTGTGCATATGGCACTCAGCTGCTTATCAGGCTATGGG	420
Db	368	CAGGCGCTCTTATGCACTAGTGTGCATATGGCACTCAGCTGCTTATCAGGCTATGGG	427
QY	421	CAGCAGCCAGCAGCCACTGCACCTTACAAGACCGCAGAGATGAAACAAGCCCACTGAGACT	480
Db	428	CAGCAGCCAGCAGCCACTGCACCTTACAAGACCGCAGAGATGAAACAAGCCCACTGAGACT	487
QY	481	ACTCAACCTCATTTGACACAGGGGGTTTACAACAGCCGACCTTAGATTTGACAGAGT	540
Db	488	ACTCAACCTCATTTGACACAGGGGGTTTACAACAGCCGACCTTAGATTTGACAGAGT	547

[illegible]



QY 1621 TGTGAAACGAGACTTGGCTGGAGAAACAGAGTGAACCAAGTGAAGCCCAAGCCT 1680  
DB 1628 TGTGAAACGAGACTTGGCTGGAGAAACAGAGTGAACCAAGTGAAGCCCAAGCCT 1687  
QY 1681 GAAGGCTTCCCTCCCGCAACCTTTCCGCCCCGGGGTGTGATCGTGGCAGAGTGGCCTT 1740  
DB 1688 GAAGGCTTCCCTCCCGCAACCTTTCCGCCCCGGGGTGTGATCGTGGCAGAGTGGCCTT 1747  
QY 1741 GGTGGCATGCGGGGAGAAAGAGTGGCTCATGATGCTGTGATCCCGGTGAATGTTTC 1800  
DB 1748 GGTGGCATGCGGGGAGAAAGAGTGGCTCATGATGCTGTGATCCCGGTGAATGTTTC 1807  
QY 1801 AGAGGTGCGGTGGTGAAGACAGAGTGGCTTCGTGTGTGGCGGGGATGGAACGAGT 1860  
DB 1808 AGAGGTGCGGTGGTGAAGACAGAGTGGCTTCGTGTGTGGCGGGGATGGAACGAGT 1867  
QY 1861 GGCCTTTGGTGAAGAAACAGAGTGGCTTCGTGTGTGGCGGGGATGGAACGAGT 1920  
DB 1868 GGCCTTTGGTGAAGAAACAGAGTGGCTTCGTGTGTGGCGGGGATGGAACGAGT 1927  
QY 1921 ATGGAGAGAAAGAGAGAGAGCTGTGAGAGACCTTGAAATGATTAAGCGGACCGCT 1980  
DB 1928 ATGGAGAGAAAGAGAGAGAGCTGTGAGAGACCTTGAAATGATTAAGCGGACCGCT 1987  
QY 1981 CAGGAGGCGAGAGATCGGCTCTAGATGAGAGACCGGACAG 2025  
DB 1988 CAGGAGGCGAGAGATCGGCTCTAGATGAGAGACCGGACAG 2032

## RESULT 2

ADSI6277  
ID ADSI6277 standard; cDNA; 2390 BP.

AC ADSI6277;

XX 18-NOV-2004 (first entry)

XX Human cDNA encoding the Ewing sarcoma protein Segid 1.

XX human; 89; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;

XX hypogonadism; androgen-resistance syndrome; testicular feminisation.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 44..2014  
/\*tag= a  
/product= "Ewing sarcoma protein"

XX EPI455190-A1.

XX 08-SEP-2004.

XX 16-FEB-2004; 2004BP-00003422.

XX 04-MAR-2003; 2003DE-01009280.

XX 25-APR-2003; 2003US-046592P.

XX (SCHD ) SCHERING AG.

XX Obendorf M, Wolf S;

XX WPI, 2004-627861/61.

XX P-PSDB; ADSI6278.

XX Determining the hormonal effects of substances, used to identify

XX pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,

XX from modulating interaction between nuclear receptors and Ewing sarcoma

XX protein.  
XX Claim 7; SEQ ID NO 1; 30p; German.  
XX This invention relates to a novel modulators that alter the interaction

CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
CC as the screening method thereof. Specifically, it refers to determining  
CC and identifying a hormonal effect brought about by test compounds that  
CC modulate either the binding of EWS to the nuclear receptor or the ligand-  
CC induced activity of this receptor. The present invention describes the  
CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
CC receptors. Accordingly, these modulators may be used in the development  
CC of pharmaceutical compositions that can diagnose and be used to treat  
CC diseases associated with receptor dysfunction such as prostatic cancer,  
CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
CC testicular feminisation. This method provides reliable, sensitive,  
CC simple, inexpensive and rapid assessment of the hormonal effects of these  
CC test compounds. This polynucleotide sequence is the cDNA encoding the  
CC human Ewing sarcoma protein of the invention.

XX Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match 100.0%; Score 2025; DB 13; Length 2390;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 2025; Conservative 0; Indels 0; Gaps 0;

QY 1 GACGACGTTGAGAGAAACAGAGAGAAAGAGAAATGCGCTCCAGATTACAGTACC 60  
DB 8 GACGACGTTGAGAGAAACAGAGAGAAAGAGAAATGCGCTCCAGATTACAGTACC 67  
QY 61 TATAGCCAGCTGAGGAGAGAGGCTACAGTGTCTTACACCGGCCAGCCCTCAAGGA 120  
DB 68 TATAGCCAGCTGAGGAGAGAGGCTACAGTGTCTTACACCGGCCAGCCCTCAAGGA 127  
QY 121 TATGACAGACCAACCAAGGATATGAGCAACAAAGCTATGAAACCTATGAGACGCCACT 180  
DB 128 TATGACAGACCAACCAAGGATATGAGCAACAAAGCTATGAAACCTATGAGACGCCACT 187  
QY 181 GATGTACGATATACCAAGGCTCAGACCACTGACCACTATGAGAGACCGCTATGCACT 240  
DB 188 GATGTACGATATACCAAGGCTCAGACCACTGACCACTATGAGAGACCGCTATGCACT 247  
QY 241 TCTTATGAGACGCTCCCACTGCTTATCTACTCCACTGCCCCCAGGATACAGCAG 300  
DB 248 TCTTATGAGACGCTCCCACTGCTTATCTACTCCACTGCCCCCAGGATACAGCAG 307  
QY 301 CCTGTCCAGGGGATATGAGCACTGTGCTTATGATACCAACCACTCTACAGTACACCAACC 360  
DB 308 CCTGTCCAGGGGATATGAGCACTGTGCTTATGATACCAACCACTCTACAGTACACCAACC 367  
QY 361 CAGGCTCTCTATGAGCTCAGTGTGATGATGAGCACTCAGGCTCTTATCCAGGCTATGAG 420  
DB 368 CAGGCTCTCTATGAGCTCAGTGTGATGATGAGCACTCAGGCTCTTATCCAGGCTATGAG 427  
QY 421 CAGCAGCAGACGACCACTGACCTACAGACCGCAGAGTGAACCAAGCCCACTGAGACT 480  
DB 428 CAGCAGCAGACGACCACTGACCTACAGACCGCAGAGTGAACCAAGCCCACTGAGACT 487  
QY 481 AGTCAACCTTAATCTAGCAACAGGGGTTACAAACCAAGCCCACTGAGTATGAGACAGT 540  
DB 488 AGTCAACCTTAATCTAGCAACAGGGGTTACAAACCAAGCCCACTGAGTATGAGACAGT 547  
QY 541 AACTCAGATTATCCCAAGGTAAGTGGAGGTAACCCATGAGCAGGCACTGACCTGCA 600  
DB 548 AACTCAGATTATCCCAAGGTAAGTGGAGGTAACCCATGAGCAGGCACTGACCTGCA 607  
QY 601 TCCTACCTCTCAACAGCTATTCCTTACACAGCCGACCTATGATGATGAGACAGTAC 660  
DB 608 TCCTACCTCTCAACAGCTATTCCTTACACAGCCGACCTATGATGATGAGACAGTAC 667  
QY 661 TCTCAGAGAAACCTATGAGGCAACGAGAGCTATGAGACAGAGTACTATGTCGA 720  
DB 668 TCTCAGAGAAACCTATGAGGCAACGAGAGCTATGAGACAGAGTACTATGTCGA 727  
QY 721 CAAGCAGCTATGAGGAGCAGCTCCCACTGATGATCCCACTGATGATGATGATGATGAT 780  
DB 728 CAAGCAGCTATGAGGAGCAGCTCCCACTGATGATCCCACTGATGATGATGATGATGAT 787

Qy	1861	GGCTTTGTTGTGAGAAAGACAGAGTGGCCCTGGGGGGGGCCCTGTGACCTTTGATGGAAACAG	1920
Db	1868	GGCTTTGTGTGAGAAACGAGGTGGCCCTGGGGGGGGCCCTGTGACCTTTGATGGAAACAG	1927
Qy	1921	ATGGAGGAAAGAAAGAGGAGACGTGGAGGACCTGGAAATAATGGATTAAAGCGACACCGT	1980
Db	1928	ATGGAGGAAAGAAAGAGGAGACGTGGAGGACCTGGAAATAATGGATTAAAGCGACACCGT	1987
Qy	1981	CAGGAGCGCAGAGATCGGCCCTTACTTAGATGCAGAGACCCCGCAG	2025
Db	1988	CAGGAGCGCAGAGATCGGCCCTTACTTAGATGCAGAGACCCCGCAG	2032

ID ABK84628 standard; cDNA; 2390 BP

AC ABK84628;

DT 14-AUG-2002 (first entry)

DE Human CDNA differentially expressed in granulocytic cells #1199.

KW Human; BS; granulocytic cell; DNA chip; bacterial infection;

KW fungal infection; sterile inflammatory disease; psoriasis;

cardiac reperfusion injury; renal reperfusion injury; ARDS;

KW Crohn's disease; ulcerative colitis; periodontal disease;

3.  $\frac{1}{2}$

XX

2000

[illegible]

**XX**

**XX**

**XX**

XX 0000 435320/45

3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527

genes associated with granulocyte activation, which serves as diagnostic

[illegible][illegible]

CC (GCA), by detecting the level of expression of gene(s) (GS) identified by

CC expression level to an expression level in an unactivated GC, where

CC modulating (M2) GA by contacting GC with an agent that alters the

capable of modulating GCA or an inflammation (especially chronic) in

pathogen or sterile inflammation using the gene expression

CC tissue, an allergic response in a subject, exposure of a subject to a

CC expression in a sample of the tissue or gene(B) from GB, where the level

CC (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject exposure of a pathogen or sterile

CC Inflammatory disease, by contacting a tissue having inflammation with a



QY 1661 GGCTTTGGTGGAGGAAGACAGAGTGAGCCCTTGAGGAGGCCCTTGACCTTTGATGAAACAG 1920  
 Db 1668 GGCCTTTGGTGGAGGAAGACAGAGTGAGCCCTTGAGGAGGCCCTTGACCTTTGATGAAACAG 1927  
 QY 1921 ATGGAGAGGAAGAGAGAGACGTGTGAGAGACCTTGAAAAATGATTAAGGCGAGCACCGT 1960  
 Db 1928 ATGGAGAGGAAGAGAGAGACGTGTGAGAGACCTTGAAAAATGATTAAGGCGAGCACCGT 1987  
 QY 1981 CAGGAGCGCAGAGATCGGCCCTTACTAGATTGACAGAGACCCCGCAG 2025  
 Db 1988 CAGGAGCGCAGAGATCGGCCCTTACTAGATTGACAGAGACCCCGCAG 2032  
 RESULT 4  
 ID ABEK64822 standard; DNA; 2390 BP.  
 AC ABEK64822;  
 DT 18-JUN-2002 (first entry)  
 DE Human benign prostatic hyperplasia gene #717.  
 KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; da.  
 OS Homo sapiens.  
 PN WO200212440-A2.  
 PD 14-FEB-2002.  
 PF 07-AUG-2001; 2001WO-US024708.  
 PR 07-AUG-2000; 2000US-0223323P.  
 PR 05-JUN-2001; 2001US-00873319.  
 XX (GENE-) GENE LOGIC INC.  
 PA (NIBS) JAPAN TOBACCO INC.  
 PI Munger WE, Kulikarni P, Getzenberg RH, Waga I, Yamamoto J;  
 DR WPI; 2002-257476/30.  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 PS  
 PS Disclosure; Page 399-400; 444pp; English.  
 XX  
 CC The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABEK4106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention

[illegible]



Db 248 TCTTATGAGACAGCTCCCACTGTTATACACTGCACTGCCCCCGAGGATACAGCAG 307  
 Qy 301 CTGTCCAGGGGATATGCACTGTGCTTATGATACCACTGCTTACATGACCAACCA 360  
 Db 308 CTTGTCCAGGGGATATGCACTGTGCTTATGATACCACTGCTTACATGACCAACCA 367  
 Qy 361 CAGGCTCTTATGAGCTCAGTGTGCACTATGAGCACTGAGCTGCTTATGAGCACTATG 420  
 Db 368 CAGGCTCTTATGAGCTCAGTGTGCACTATGAGCACTGAGCTGCTTATGAGCACTATG 427  
 Qy 421 CAGGAGCAGAGCAGCAGTGCATCTTACAGACCGAGAGATGGAACCAAGCCCACTGAGCT 480  
 Db 428 CAGAGCAGAGCAGCAGTGCATCTTACAGACCGAGAGATGGAACCAAGCCCACTGAGCT 487  
 Qy 481 AGTCACTCTTATGAGCACTGAGGGGTTTCAACCAAGCCCACTGAGATATGACAGAGT 540  
 Db 488 AGTCACTCTTATGAGCACTGAGGGGTTTCAACCAAGCCCACTGAGATATGACAGAGT 547  
 Qy 541 AACTACAGTTATCCCAAGTACCTGAGGAGTACCCCACTGAGCAGCAGTCACTGACCTCA 600  
 Db 548 AACTACAGTTATCCCAAGTACCTGAGGAGTACCCCACTGAGCAGCAGTCACTGACCTCA 607  
 Qy 601 TCTTACCTCTTACAGCTATTTCTCTTACAGCAGCAGCTAGTTATGATCAGAGAGTTAC 660  
 Db 608 TCTTACCTCTTACAGCTATTTCTCTTACAGCAGCAGCTAGTTATGATCAGAGAGTTAC 667  
 Qy 661 TCTTACAGCAACCTTATGAGCACTGAGGAGTACCAAGCAGTATGAGCAGAGTATGTCGA 720  
 Db 668 TCTTACAGCAACCTTATGAGCACTGAGGAGTACCAAGCAGTATGAGCAGAGTATGTCGA 727  
 Qy 721 CAAAGCAGCTATGAGGAGCAGCCTCCCACTAGTTACCAACCCCAATGATCTTACAGC 780  
 Db 728 CAAAGCAGCTATGAGGAGCAGCCTCCCACTAGTTACCAACCCCAATGATCTTACAGC 787  
 Qy 781 CAACTCCAACTCAATATAGCCAAAGAGCAGCAGCTACGAGGAGCAGAGTTCCATCCGA 840  
 Db 788 CAACTCCAACTCAATATAGCCAAAGAGCAGCAGCTACGAGGAGCAGAGTTCCATCCGA 847  
 Qy 841 CAGAGCAACCCCACTATGAGGAGTATGAGGAGAGTGTGAGAGATTTTCCGAGCA 900  
 Db 848 CAGAGCAACCCCACTATGAGGAGTATGAGGAGAGTGTGAGAGATTTTCCGAGCA 907  
 Qy 901 GAGAGAGACCGGAGCAGTATGAGGAGTATGAGGAGAGTGTGAGAGATTTTCCGAGCA 960  
 Db 908 GAGAGAGACCGGAGCAGTATGAGGAGTATGAGGAGAGTGTGAGAGATTTTCCGAGCA 967  
 Qy 961 GGTGAGGAGCAGTATGAGGAGTATGAGGAGAGAGAGAGAGTGTGAGATTTTCCGAGCA 1020  
 Db 968 GGTGAGGAGCAGTATGAGGAGTATGAGGAGAGAGAGAGAGTGTGAGATTTTCCGAGCA 1027  
 Qy 1021 GAGGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGAGAGTGTGAGATTTTCCGAGCA 1080  
 Db 1028 GAGGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGAGAGTGTGAGATTTTCCGAGCA 1087  
 Qy 1081 GAGGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGAGAGTGTGAGATTTTCCGAGCA 1140  
 Db 1088 GAGGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGAGAGTGTGAGATTTTCCGAGCA 1147  
 Qy 1141 AATGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 Db 1148 AATGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1207  
 Qy 1201 AATGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 Db 1208 AATGACAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1267  
 Qy 1261 CCAAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 Db 1268 CCAAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1327  
 Qy 1321 TGGTTTATGAGGAGATTTTCAAGGAGCAACTTAAAGTCTCTTGTCTGAGAGAG 1380

Db 1328 TGGTTTATGAGGAGATTTTCAAGGAGCAACTTAAAGTCTCTTGTCTGAGAGAG 1387  
 Qy 1381 CTTCCAAATGAAAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1440  
 Db 1388 CTTCCAAATGAAAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1447  
 Qy 1441 CCACTCCGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1500  
 Db 1448 CCACTCCGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1507  
 Qy 1501 GGCCTGAGAGAGATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1560  
 Db 1508 GGCCTGAGAGAGATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1567  
 Qy 1561 CCCTCTGAGAGAGATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1620  
 Db 1568 CCCTCTGAGAGAGATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAGGAGT 1627  
 Qy 1621 TGTGAAACCAAGATCTTCCGCTGAGAGACAGAGTGCACCAAGTGTAAAGGCCCAAGCT 1680  
 Db 1628 TGTGAAACCAAGATCTTCCGCTGAGAGACAGAGTGCACCAAGTGTAAAGGCCCAAGCT 1687  
 Qy 1681 GAAAGCTTCTCTCCGCACTCTTCCGCTGAGAGACAGAGTGCACCAAGTGTAAAGGCCCAAGCT 1740  
 Db 1688 GAAAGCTTCTCTCCGCACTCTTCCGCTGAGAGACAGAGTGCACCAAGTGTAAAGGCCCAAGCT 1747  
 Qy 1741 GGTGCAATGAGGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1800  
 Db 1748 GGTGCAATGAGGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1807  
 Qy 1801 AGAGTGTGAGGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1860  
 Db 1808 AGAGTGTGAGGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1867  
 Qy 1861 GGTGTTGTGAGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1920  
 Db 1868 GGTGTTGTGAGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1927  
 Qy 1921 ATGAGAGAGAGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1980  
 Db 1928 ATGAGAGAGAGAGAGAGAGTGCCTTCAAGATGAGTGTGAGTGTGAGTGTGAGTGTG 1987  
 Qy 1981 CAGAGGCGAGAGATGCGCTTCTAGATGACAGAGACCCCGCAGA 2025  
 Db 1988 CAGAGGCGAGAGATGCGCTTCTAGATGACAGAGACCCCGCAGA 2032

RESULT 6  
 ACN40903  
 ID ACN40903 standard; cDNA; 2390 BP.  
 XX  
 AC ACN40903;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.  
 XX  
 KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KW central nervous system cancer; bladder cancer; pancreatic cancer;  
 KW cervical cancer; melanoma; leukemia; hybridization probe;  
 KW chromosome identification; chromosome mapping; gene mapping;  
 KW gene therapy; cytostatic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02004030615-A2.  
 XX  
 PD 15-APR-2004.  
 XX  
 PF 29-SEP-2003; 2003WO-US028547.  
 XX







QY	1561	CCCTCTGAGGAGAAAAGTCCACACCGAGCTGAGAGTGGCAGTGTCCCAATCCGGGT	1620
Db	1568	CCCTCTGAGGAGGAAAGTCCACACCGAGCTGAGAGTGGCAGTGTCCCAATCCGGGT	1627
QY	1621	TGTGGAAACCAAGACTTGGCTGGAGAAAGAGTGTCAACAGTGTAAAGGCCCCAAAGCTT	1680
Db	1628	TGTGGAAACCAAGACTTGGCTGGAGAAAGAGTGTCAACAGTGTAAAGGCCCCAAAGCTT	1687
QY	1681	GAAGGCTTCCCTCCGCGCAACCTTTCCGCGCCCGGGGTGGTGAATCGTGGCAGAGTGGCCCT	1740
Db	1688	GAAGGCTTCCCTCCGCGCAACCTTTCCGCGCCCGGGGTGGTGAATCGTGGCAGAGTGGCCCT	1747
QY	1741	GGTGGCAATGCGGGAGAGAAAGAGTGGCTTCAATGAAATCGTGTGGTCCCGGTGAATGTTT	1800
Db	1748	GGTGGCAATGCGGGAGAGAAAGAGTGGCTTCAATGAAATCGTGTGGTCCCGGTGAATGTTT	1807
QY	1801	AGAGTGGCCCTGTGTGAGACAGAGTGGCTTCCGTGTGTGACCCGGGGACATGAAACGAGGT	1866
Db	1808	AGAGTGGCCCTGTGTGAGACAGAGTGGCTTCCGTGTGTGACCCGGGGACATGAAACGAGGT	1867
QY	1861	GGCTTTGGTGTGAGGAAAGAAGAGTGGTGGCCCTGTGGGGGCCCTCTGACCTTTGATGGAAAG	1920
Db	1868	GGCTTTGGTGTGAGGAAAGAAGAGTGGTGGCCCTGTGGGGGCCCTCTGACCTTTGATGGAAAG	1927
QY	1921	ATGGAGGAAAGAAAGAGAGACGTGTGAGAGACTTGGAAAAATGATTAAGGCGAGACCTGT	1988
Db	1928	ATGGAGGAAAGAAAGAGAGACGTGTGAGAGACTTGGAAAAATGATTAAGGCGAGACCTGT	1987
QY	1981	CAGAGCGCAGAGATCGGCGCTTACTAGATGACAGAGACCCCGCAGA	2025
Db	1988	CAGAGCGCAGAGATCGGCGCTTACTAGATGACAGAGACCCCGCAGA	2032

RESULT 7	
AAQ50643	
ID	AAQ50643 standard; cDNA; 2371 BP.
XX	
AC	AAQ50643;
XX	
DT	25-MAR-2003 (revised)
DT	26-MAY-1994 (first entry)
DE	Human Ews gene clone BFLAC5 from foetal brain cDNA library.
XX	
XX	chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;
KW	malignant melanoma; hum-fil-1;
KW	primitive peripheral neuroectodermal tumour; human chromosome 11;
KW	human chromosome 22; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	25..1995
FT	location/Qualifiers
FT	/*tag= a
FT	/product= "EWS_protein"
FT	/tranyl_except= pos:1729..1731; aa:Val
FT	2143..2148
FT	/*tag= b
FT	2331..2336
FT	polyA_signal
FT	/*tag= c
XX	
XX	MO9323549-A2.
PN	
XX	
PD	25-NOV-1993.
XX	
PP	19-MAY-1993; 93WO-FR000494.
XX	
PR	20-MAY-1992; 92FR-00006123.
XX	
PA	(CNRS ) CNRS CENT NAT RECH SCT.
PI	Aurias A, Delattre O, Deamaze C, Melot T, Peter M, Plougastel B;

Pt	Thomas G; Zucman J;
Xx	WP1; 1993-386580/48.
Dn	P-PsDB; AAR#4555.
Xx	New nucleic acid of EMS gene and its hybrid(s) - contg. gene sequence
Pt	involved in chromosomal trans-location, also derived mRNA, probes, fusion
Pt	proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
Xs	Disclosure; Fig 6; 123pp; French.
Cc	The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA
Cc	library (Stratagene cat.# 936706). The clone BFL1C5 was identified and
Cc	sequenced. It represents the entire coding region and 3'-UTR of the Ews
Cc	gene. (Updated on 25-MAR-2003 to correct PN field.)
Sq	Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;
Qy	Query Match            99.3%; Score 2011.4; DB 2; Length 2371;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2012; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Dn	13 GAGAACGAGGAGGAAGAAGAAAATGGCGCTCCACGATTACAGTACTTAAGCAAGCT 72
Oy	1 GAGAAACGAGGAGGAAGAAGAAAATGGCGCTCCACGATTACAGTACTTAAGCAAGCT 60
Dn	73 GCACGCCGACGAGGGGCTACAGTGCTTTACAACCGCCCAAGCCCACTCAAAGATATGCACAGC 132
Oy	61 GCACGCCGACGAGGGGCTACAGTGCTTTACAACCGCCCAAGCCCACTCAAAGATATGCACAGC 120
Oy	133 ACCGAGCATATGSGCAACAAAGCTATGSAACCTATGACACAGCCCACTGATGCAGCTAT 192
Dn	121 ACCGAGCATATGSGCAACAAAGCTATGSAACCTATGACACAGCCCACTGATGCAGCTAT 180
Oy	193 ACCGAGGTCAAGCACTGCAACCTATGSGCAGACCAGCTATGCAACTTCTTAGACAG 252
Dn	181 ACCGAGGTCAAGCACTGCAACCTATGSGCAGACCAGCTATGCAACTTCTTAGACAG 240
Oy	253 CCTCCCACGTGTTATATCTAATCTCCAATGCCCCCAAGGATACAGCACGCTGTCCAGGG 312
Dn	241 CCTCCCACGTGTTATATCTAATCTCCAATGCCCCCAAGGATACAGCACGCTGTCCAGGG 300
Oy	313 TATGGCACTGGTGTCTTATATGATACCAACCACTGCTCAAGTCACCAACCAAGGCTCTTAT 372
Dn	301 TATGGCACTGGTGTCTTATATGATACCAACCACTGCTCAAGTCACCAACCAAGGCTCTTAT 360
Oy	373 GCAGTCAGTGTGATATATGGAAGCTCAGGCTGCTTATCCAGCTATGSGCAGACGCACGA 432
Dn	361 GCAGTCAGTGTGATATATGGAAGCTCAGGCTGCTTATCCAGCTATGSGCAGACGCACGA 420
Oy	433 GCCACTGACCTTACAAGACCGCAGATGGAACAAAGCCCACTGAGACTATGATCAACTCA 492
Dn	421 GCCACTGACCTTACAAGACCGCAGATGGAACAAAGCCCACTGAGACTATGATCAACTCA 480
Oy	493 TCTTGCAACAGGGGGTTTACAACGAGCCGCTTAGAGATATGGAACAGAGTAACTACAGTTAT 552
Dn	481 TCTTGCAACAGGGGGTTTACAACGAGCCGCTTAGAGATATGGAACAGAGTAACTACAGTTAT 540
Oy	553 CCCGAGTACCTGGAGAGTACCCCAATGACAGCAGTCACTGCACTCAATCCTACCTCCT 612
Dn	541 CCCGAGTACCTGGAGAGTACCCCAATGACAGCAGTCACTGCACTCAATCCTACCTCCT 600
Oy	613 ACCAGCTATTCCTCTTACACAGCCCACTATGATATGATCAAGACAGTTACTCTCAGACAAAC 672
Dn	601 ACCAGCTATTCCTCTTACACAGCCCACTATGATATGATCAAGACAGTTACTCTCAGACAAAC 660
Oy	673 ACCTATGGGCAACGAGAGCTATATGAAAGAGAGTACTATGTCACAAAGACAGCTAT 732
Dn	661 ACCTATGGGCAACGAGAGCTATATGAAAGAGAGTACTATGTCACAAAGACAGCTAT 720
Oy	733 GGGGACAGAGCTCCCACTAGTTATCCCAACCCAACTGATCCTACAGCAAGACTCCAGT 792
Dn	721 GGGGACAGAGCTCCCACTAGTTATCCCAACCCAACTGATCCTACAGCAAGACTCCAGT 780

OY	793	CAATTATGACAA	CAGAAGACAGCTA	CCGGGACGAGACTT	CAATTC	CCGACAGACCA	CAACCC	852
Db	791	CAATTATGACAA	CAGAAGACAGCTA	CCGGGACGAGACTT	CAATTC	CCGACAGACCA	CAACCC	840
OY	853	AGTAGCATGG	GTATTATGG	CAGAGTCT	GAGAGATTTT	CCGACAGAGAAC	CCGG	912
Db	841	AGTAGCATGG	GTATTATGG	CAGAGTCT	GAGAGATTTT	CCGACAGAGAAC	CCGG	900
OY	913	AGCATGATGG	CCCTGATPAC	CCGGGGCAGGGG	AAGAGGGGATTT	GATTCGTGAGGCATG	972	
Db	901	AGCATGATGG	CCCTGATPAC	CCGGGGCAGGGG	AAGAGGGGATTT	GATTCGTGAGGCATG	960	
OY	973	AGCAGAGGTGG	CGCGGAGAGAG	ACCGGTGGAT	TGGCAGACGCT	GTGAGACGAGGTGGC	1032	
Db	961	AGCAGAGGTGG	CGCGGAGAGAG	ACCGGTGGAT	TGGCAGACGCT	GTGAGACGAGGTGGC	1020	
OY	1033	TTCAATTAAG	CCCTGTGGAC	CCCATGGATG	AGACCAAGTCTT	GATCTAGGCCCATCTGTA	1092	
Db	1021	TTCAATTAAG	CCCTGTGGAG	CCCATGGATG	AGACCAAGTCTT	GATCTAGGCCCATCTGTA	1080	
OY	1093	GATCCAGATGA	AGACTCTGAC	CAACAGTGCAT	TTTATGTAC	AAAGATTTAATGACAGTGTG	1152	
Db	1081	GATCCAGATGA	AGACTCTGAC	CAACAGTGCAT	TTTATGTAC	AAAGATTTAATGACAGTGTG	1140	
OY	1153	ACTTAGATGAT	CTTGGCAGACTT	CTTTAAGCAGTGTGG	GTGTTAAGATGAA	CAAGAGA	1212	
Db	1141	ACTTAGATGAT	CTTGGCAGACTT	CTTTAAGCAGTGTGG	GTGTTAAGATGAA	CAAGAGA	1200	
OY	1213	ACTGGGCAAC	CCCATGATTC	CATCTACCTG	GCAGAGAA	CAAGAAAGCCCAAGGCAT	1272	
Db	1201	ACTGGGCAAC	CCCATGATTC	CATCTACCTG	GCAGAGAA	CAAGAAAGCCCAAGGCAT	1260	
OY	1273	GCCAAGTGTCT	TATGAAGACCA	CCCATGCTGC	CAAGCTGCGTGG	ATGTTGATGTGG	1332	
Db	1261	GCCAAGTGTCT	TATGAAGACCA	CCCATGCTGC	CAAGCTGCGTGG	ATGTTGATGTGG	1320	
OY	1333	AAAGATTTT	CAAGGAGCAAA	CTTAAATCTT	CCCTTGTG	CGAAGAAAGCTTCAATGAC	1392	
Db	1321	AAAGATTTT	CAAGGAGCAAA	CTTAAATCTT	CCCTTGTG	CGAAGAAAGCTTCAATGAC	1380	
OY	1393	AGTATGCGGG	TGCTGTGCA	CCCCGTGAGGG	CAAGGCATG	CCACACATCCTCGTGG	1452	
Db	1381	AGTATGCGGG	TGCTGTGCA	CCCCGTGAGGG	CAAGGCATG	CCACACATCCTCGTGG	1440	
OY	1453	GCTCCAGAGG	CCCGAGAGGTCT	TGTGGGGA	CCCATGTGCT	GCATGTGAGGCGCTGAGGA	1512	
Db	1441	GCTCCAGAGG	CCCGAGAGGTCT	TGTGGGGA	CCCATGTGCT	GCATGTGAGGCGCTGAGGA	1500	
OY	1513	GATAGAGAGG	CTTCTCTT	CCAAAGAGAAC	CCCCGGGTTCT	CCGAGGGAAC	CCCTCTGAGGA	1572
Db	1501	GATAGAGAGG	CTTCTCTT	CCAAAGAGAAC	CCCCGGGTTCT	CCGAGGGAAC	CCCTCTGAGGA	1560
OY	1573	GGAAGACGT	CCACGACCCG	AGCTGAGAC	CTGGCAGTGT	CCCAATCCGGGTTGTG	GAAGAACAG	1632
Db	1561	GGAAGACGT	CCACGACCCG	AGCTGAGAC	CTGGCAGTGT	CCCAATCCGGGTTGTG	GAAGAACAG	1620
OY	1633	AACCTTCCG	CTGAGAGAC	CAGAGTGCA	CCAGTGTAA	AGCCCCAAAGCTTGAAGGCTTCTC	1692	
Db	1621	AACCTTCCG	CTGAGAGAC	CAGAGTGCA	CCAGTGTAA	AGCCCCAAAGCTTGAAGGCTTCTC	1680	
OY	1693	CCGCAAC	CCCTTCCG	CCCCCGGGG	TGATGATCTG	TGGCAGAGGTGGCC	CTGTGTGGCATG	1752
Db	1681	CCGCAAC	CCCTTCCG	CCCCCGGGG	TGATGATCTG	TGGCAGAGGTGGCC	CTGTGTGGCATG	1740
OY	1753	GGAGAGAA	GAGTGGCT	CTCATG	GAATCTGTGTGT	GTCCCGGTGGAAT	GTTCAGAGGTGGCGT	1812
Db	1741	GGAGAGAA	GAGTGGCT	CTCATG	GAATCTGTGTGT	GTCCCGGTGGAAT	GTTCAGAGGTGGCGT	1800
OY	1813	GGTGGAG	CAAGAGTGT	CTTCCGTGTGT	GGCCGGG	CAATGAC	CCGAGGTGGCTTTGTGTGGA	1872
Db	1801	GGTGGAG	CAAGAGTGT	CTTCCGTGTGT	GGCCGGG	CAATGAC	CCGAGGTGGCTTTGTGTGGA	1860

QY	1873	GAAGACGAGGTTGCGCCCTGGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGAGA	1932
QY	1873	GAAGACGAGGTTGCGCCCTGGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGAGA	1932
Db	1861	GGAAAGACGAGGTTGCGCCCTGGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGAGA	1920
QY	1933	AGAGGAGGACCTGAGGAGCCTTGAGAAAAATGATTAAGGCGGAGCA	1992
Db	1921	AGAGGAGGACCTGAGGAGCCTTGAGAAAAATGATTAAGGCGGAGCACCGTCAGAGGCGCAGA	1980
QY	1993	GATCGGCCCTTACTAGATGCAGAGAGACCCCGCAGA	2025
Db	1981	GATCGGCCCTTACTAGATGCAGAGAGACCCCGCAGA	2013

## RESULT 8

AAS62262/c  
ID AAS62262 standard; cDNA; 2176 BP

AC AAS62262

DT 14-FEB-2002 (first entry)

CDNA sequence #49 encoding novel human secreted protein.

Human secreted protein; hyperproliferative disorder; autoimmune disorder;

KW infectious disorder; gene therapy; antimicrobial; hepatotropic;

388

**XX**

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

211

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

XX  
XX  
(continued)

**XX**

PI Gulukota K, Graham JR;  
.....

DR WPI; 2002-010900/01

PT New polynucleotides encoding se

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

3X3

CC which encode human secreted proteins. The cDNA sequences have been

method for producing proteins from these polynucleotide sequences. The

CC activity and production, and the cell is also useful for identifying

CC encoding the secreted proteins. The sequences of the invention are used for encoding the secreted proteins.

CC Immune deficiency disorders (e.g. severe combined immunodeficiency (SCID), autoimmune disorders (e.g. multiple sclerosis), blood disorders

CC (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and infectious disorders (e.g. borreliosis). The following table summarizes

CC the invention are also useful in gene therapy. AAS62214-AAS62838  
CC relevant to the DNA sequences of the invention that encode for novel

CC secreted proteins  
xy

Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Query Match	98.14;	Score 1986.4;	DB 6;	Length 2176;
Post Local Similarity	99.98;	Score No. 0;		

Matches 1998; Conservative 0; Mismatches 1; Indels 1; Gaps

QY 26 AAGAGAGAAATGCGCTCCACGGATTAAGTACTTATAGCAAGCTGACGGCAGAGG 85  
DB 2171 AAGGTGAGAAATGCGCTCCACGGATTAAGTACTTATAGCAAGCTGACGGCAGAGG 2112  
QY 86 GCTACAGTGTCTTACCGCCAGCCCACTCAAGGATATGCAAGACCAACCCAGGATATG 145  
DB 2111 GCTACAGTGTCTTACCGCCAGCCCACTCAAGGATATGCAAGACCAACCCAGGATATG 2052  
QY 146 GGCACAAAGCTATGGAACCTATGACAGCCCACTGATGTCAAGTATACCAAGGCTTACA 205  
DB 2051 GGCACAAAGCTATGGAACCTATGACAGCCCACTGATGTCAAGTATACCAAGGCTTACA 1992  
QY 206 CCACTGCAACCTATGAGGAGAGCGGCTATGCACTTCTTATGACAGCTCCCACTGAT 265  
DB 1991 CCACTGCAACCTATGAGGAGAGCGGCTATGCACTTCTTATGACAGCTCCCACTGAT 1932  
QY 266 ATACTACTCCACTGCCCCCGGACATACAGCCAGCTGTCCAGGGGATATGCACTGAT 325  
DB 1931 ATACTACTCCACTGCCCCCGGACATACAGCCAGCTGTCCAGGGGATATGCACTGAT 1872  
QY 326 CTTATGATTAACCACTGCTTACAGTCAACCAACCCAGGCTCTTATGACAGTCAATCTG 385  
DB 1871 CTTATGATTAACCACTGCTTACAGTCAACCAACCCAGGCTCTTATGACAGTCAATCTG 1812  
QY 386 CATATGCACTCAAGCTGCTTATCAGGCTATGAGGAGAGCCAGCAGCCACTGCACTTA 445  
DB 1811 CATATGCACTCAAGCTGCTTATCAGGCTATGAGGAGAGCCAGCAGCCACTGCACTTA 1752  
QY 446 CAAGACCGCAGATGAGAAACAGGCCCACTGAGTACTCAACTCAATCTAGCAGAGGG 505  
DB 1751 CAAGACCGCAGATGAGAAACAGGCCCACTGAGTACTCAACTCAATCTAGCAGAGGG 1693  
QY 506 GTTACAAACAGCCAGCTTATGAGATATGAGACAGATTAATCAAGTATACCCAGGTAAGT 565  
DB 1692 GTTACAAACAGCCAGCTTATGAGATATGAGACAGATTAATCAAGTATACCCAGGTAAGT 1633  
QY 566 GGAAGCTACCCCATGACAGTCACTGCACTCCATCTTACCTCTACAGGATATCT 625  
DB 1632 GGAAGCTACCCCATGACAGTCACTGCACTCCATCTTACCTCTACAGGATATCT 1573  
QY 626 CTACACAGCCGACTAGTATGATGATGACAGCACTTACTCTCAGCAGAAACCTATGAGGAG 685  
DB 1572 CTACACAGCCGACTAGTATGATGATGACAGCACTTACTCTCAGCAGAAACCTATGAGGAG 1513  
QY 686 CGAGCACTATGACAGCAGAGATGATGATGATCAACAAAGAGATATGAGGAGAGGCTC 745  
DB 1512 CGAGCACTATGACAGCAGAGATGATGATGATCAACAAAGAGATATGAGGAGAGGCTC 1453  
QY 746 CCACTAGTACCAACCCCAACTGATCTTACAGCAGCAAGCTCAAGTCAATATATGCAAC 805  
DB 1452 CCACTAGTACCAACCCCAACTGATCTTACAGCAGCAAGCTCAAGTCAATATATGCAAC 1393  
QY 806 AGAGCAGCAGCTAGGAGAGAGATTCATTCAGACAGAACCAACCCAGTATGAGGAG 865  
DB 1392 AGAGCAGCAGCTAGGAGAGAGATTCATTCAGACAGAACCAACCCAGTATGAGGAG 1333  
QY 866 TTTATGAGCAGAGATCTGAGAGATTTTCCGGAACAAGAGAGAACCCGAGCAGTATGAGG 925  
DB 1332 TTTATGAGCAGAGATCTGAGAGATTTTCCGGAACAAGAGAGAACCCGAGCAGTATGAGG 1273  
QY 926 CTGATTAACCGGAGAGAGAGAGAGGAGATTTGATCTGAGAGCAGTATGAGCAGAGGAG 985  
DB 1272 CTGATTAACCGGAGAGAGAGAGAGGAGATTTGATCTGAGAGCAGTATGAGCAGAGGAG 1213  
QY 986 GGGAGAGAGAGAGAGAGAGAGAGAGAGGAGATTCGAGAGAGAGAGGAGTTCATATAGCTG 1045  
DB 1212 GGGAGAGAGAGAGAGAGAGAGAGAGAGGAGATTCGAGAGAGAGAGGAGTTCATATAGCTG 1153  
QY 1046 GTGAGACCATGATGAGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCAATATAG 1105  
DB 1152 GTGAGACCATGATGAGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCAATATAG 1093  
QY 1106 ACTGTGACAAAGTGCATTTATGATCAAGAGATTAATGACAGTGTGACTCTAGATGATC 1165

DB 1092 ACTGTGACAAAGTGCATTTATGATCAAGAGATTAATGACAGTGTGACTCTAGATGATC 1033  
QY 1166 TGGCAGACTCTTTTAAAGCAGTGTGGGTTGTTAAGTGAACAAAGAACTGGGCAACCA 1225  
DB 1032 TGGCAGACTCTTTTAAAGCAGTGTGGGTTGTTAAGTGAACAAAGAACTGGGCAACCA 973  
QY 1226 TGAATCAACCTTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCT 1285  
DB 972 TGAATCAACCTTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCT 913  
QY 1286 ATGAAAGCCCAACCACTGCAAGAGCTGCGGTGGAATGTTTATGAGAAAGATTTTCAAG 1345  
DB 912 ATGAAAGCCCAACCACTGCAAGAGCTGCGGTGGAATGTTTATGAGAAAGATTTTCAAG 853  
QY 1346 GAGCAAACTTAAAGTCTCCCTTGTCTGCAAGAAAGCTCCATATGAACAGTATCGGGGTG 1405  
DB 852 GAGCAAACTTAAAGTCTCCCTTGTCTGCAAGAAAGCTCCATATGAACAGTATCGGGGTG 793  
QY 1406 GTCTGCACCCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCC 1465  
DB 792 GTCTGCACCCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCC 733  
QY 1466 CAGAGAGTCTGAGGAGACCCATGAGTTCGATGAGAGGCGGTGAGAGATAGAGAGCT 1525  
DB 732 CAGAGAGTCTGAGGAGACCCATGAGTTCGATGAGAGGCGGTGAGAGATAGAGAGCT 673  
QY 1526 TCCCTCCAGAGAGACCCCGGGGTTCCCGAGAGAAACCTTGTGAGAGAGAAAGTCCAGC 1585  
DB 672 TCCCTCCAGAGAGACCCCGGGGTTCCCGAGAGAAACCTTGTGAGAGAGAAAGTCCAGC 613  
QY 1586 ACCGAGCTGAGAGCTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTGCGCTGGA 1645  
DB 612 ACCGAGCTGAGAGCTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTGCGCTGGA 553  
QY 1646 GAAAGAGTGCACCAAGTGAAGGCCCAAGCTTGAAGGCTTCTCCGCAACCTTTC 1705  
DB 552 GAAAGAGTGCACCAAGTGAAGGCCCAAGCTTGAAGGCTTCTCCGCAACCTTTC 493  
QY 1706 CGCCCCCGGAGTGTATCTGAGAGAGTGGCCCTGTGCAATCGCGGAGAGAAAGAGTG 1765  
DB 492 CGCCCCCGGAGTGTATCTGAGAGAGTGGCCCTGTGCAATCGCGGAGAGAAAGAGTG 433  
QY 1766 GCCCTAATGATCGTGTGAGTCCCGTGAAGTGTTCAGAGGAGTGGCCGTGAGAGACAGAG 1825  
DB 432 GCCCTAATGATCGTGTGAGTCCCGTGAAGTGTTCAGAGGAGTGGCCGTGAGAGACAGAG 373  
QY 1826 GTGAGCTTCCGTGTGAGCCGAGGAGATGGAACGAGGTGCTTGTGAGAGAAAGAGAGTG 1885  
DB 372 GTGAGCTTCCGTGTGAGCCGAGGAGATGGAACGAGGTGCTTGTGAGAGAAAGAGAGTG 313  
QY 1886 GCCCTGAGGAGCCCTCTGAGACTTTGATGAGAACAGATGAGAGAAAGAGAGAGAGTG 1945  
DB 312 GCCCTGAGGAGCCCTCTGAGACTTTGATGAGAACAGATGAGAGAAAGAGAGAGAGTG 253  
QY 1946 GAGGACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTACT 2005  
DB 252 GAGGACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTACT 193  
QY 2006 AGATGCAAGAGACCCGAGAG 2025  
DB 192 AGATGCAAGAGACCCGAGAG 173

RESULT 9  
AAS70647  
ID AAS70647 standard; cDNA; 2388 BP.  
XX AAS70647;  
AC  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #6451.



Db	1339	GTGGAAATGGTTTGAATGGAAAGATTTTTCAGAGGAGCAAACTTAAATGTCCTCTTGCTCGG	1398
Qy	1375	AAGAAGCTCCAAATGAACATATATCGGGGTGTCTGCCACCCCGTAGGGCAGAGGCATG	1434
Db	1399	AAGAAGCTCTCAATGAACATATATCGGGGTGTCTGCCACCCCGTAGGGCAGAGGCATG	1458
Qy	1435	CCACCAACAATCTCGGTGAGGTCCTCAGAGAGCCCAAGAGGTCTTGAGGGAGACCAATGGGTGCG	1494
Db	1459	CCACCAACAATCTCGGTGAGGTCCTCAGAGAGCCCAAGAGGTCTTGAGGGAGACCAATGGGTGCG	1518
Qy	1495	ATGGAGAGCCGTGAGAGATAGAGAGGCTTCTCCTCCAAAGAGACCCCGGGGTTCCCGA	1554
Db	1519	ATGGAGAGCCGTGAGAGATAGAGAGGCTTCTCCTCCAAAGAGACCCCGGGGTTCCCGA	1578
Qy	1555	GGGAACCCCTGTGAGAGAGAAACGTCCAGACCCGAGCTGAGACGTGCACTGTCCAAAT	1614
Db	1579	GGGAACCCCTGTGAGAGAGAAACGTCCAGACCCGAGCTGAGACGTGCACTGTCCAAAT	1638
Qy	1615	CCGGGTTGTGAAACCAACAATCTTGCGCTCTGAGAAACAAGTGTCAACCAATGTAAAGCCCCA	1674
Db	1639	CCGGGTTGTGAAACCAACAATCTTGCGCTCTGAGAAACAAGTGTCAACCAATGTAAAGCCCCA	1698
Qy	1675	AAGCCTGAAGGCTTCTCCGCGCAACCTTTCGCGCCCGGGGTGTGATCGTGGCAGAGGT	1734
Db	1699	AAGCCTGAAGGCTTCTCCGCGCAACCTTTCGCGCCCGGGGTGTGATCGTGGCAGAGGT	1758
Qy	1735	GGCCCTGTGTGCAATCGCGGAGAGAAAGAGGTGGCTCATGATCGTGTGTCTCCCGTGTGA	1794
Db	1759	GGCCCTGTGTGCAATCGCGGAGAGAAAGAGGTGGCTCATGATCGTGTGTCTCCCGTGTGA	1818
Qy	1795	ATGTTCAAGTGTGCGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGCGGGGGCATAGAC	1854
Db	1819	ATGTTCAAGTGTGCGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGCGGGGGCATAGAC	1878
Qy	1855	CGAGGTGCTTGTGTGAGAGAAAGACAGAGTGGCCCTGTGGGGGACCCCTGTGACCTTTGATG	1914
Db	1879	CGAGGTGCTTGTGTGAGAGAAAGACAGAGTGGCCCTGTGGGGGACCCCTGTGACCTTTGATG	1938
Qy	1915	GAACAGATGGAGGAAAGAAAGAGAGACGTGAGAGACCTGGAATAATGATTAAGGCGAG	1974
Db	1939	GAACAGATGGAGGAAAGAAAGAGAGACGTGAGAGACCTGGAATAATGATTAAGGCGAG	1998
Qy	1975	CACGCTCAGAGAGCGCAGAGATCGGCGCTTATGATGCAAGAACCCCGCAGA	2025
Db	1999	CACGCTCAGAGAGCGCAGAGATCGGCGCTTATGATGCAAGAACCCCGCAGA	2049

RESULT 10

ACN37537

ID 1D ACN37537 standard; cDNA, 2177 BP.

ACN37537;

DT 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) cDNA DNA233971, SEQ ID NO:490.

XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;

XX tumour; diagnosis; cell proliferative disorder; breast cancer;

KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;

KM central nervous system cancer; bladder cancer; pancreatic cancer;

KM cervical cancer; melanoma; leukaemia; hybridisation probe;

KM chromosome identification; chromosome mapping; gene mapping;

XX gene therapy; cytostatic; gene; ss.

OS Homo sapiens.

XX

XX WO2004030615-A2.

XX

PD 15-APR-2004.

XX

PF 29-SEP-2003; 2003WO-US028547.

XX	02-0CT-2002; 2002US-0414971P.	
XX		
XX	(GETH ) GENENTECH INC.	
XX		
XX	Wu TD, Zhang Z, Zhou Y;	
XX		
XX	WI: 2004-347921/32.	
XX	P-PSDB: ABM80197.	
XX		
XX	New tumor-associated antigenic target polypeptides and nucleic acids,	
XX	useful in preparing a medicament for treating or detecting a	
XX	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
XX	prostate cancer or tumor.	
XX		
XX	Claim 1; SEQ ID NO 490; 7273bp; English.	
XX		
XX	The invention relates to human tumour-associated antigenic target (TAT)	
XX	polypeptides, and their related nucleic acids. The TAT polypeptides are	
XX	overexpressed in cancer tissues compared to normal tissues, and may thus	
XX	serve as effective targets for the diagnosis and treatment of cancer in	
XX	mammals. The invention also relates to nucleic acid and polypeptide	
XX	sequences at least 80% identical to the TAT nucleic acids and	
XX	polypeptides; expression vectors and host cells comprising a TAT nucleic	
XX	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
XX	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
XX	TAT polypeptide; and methods and compositions for the treatment or	
XX	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
XX	antibodies, antagonists, binding molecules and compositions are useful	
XX	for diagnosing or treating a cell proliferative disorder associated with	
XX	increased TAT expression, particularly cancers such as breast cancer,	
XX	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder	
XX	cancer, pancreatic cancer, cervical cancer, cancers of the central	
XX	nervous system, melanoma and leukaemia. TAT nucleic acids may further be	
XX	used as hybridisation probes, in chromosome and gene mapping, in	
XX	chromosome identification and in gene therapy. The present sequence	
XX	represents a TAT nucleic acid of the invention	
XX		
XX	Sequence 2177 BP; 602 A; 539 C; 605 G; 431 T; 0 U; 0 Other;	
XX		
XX	Query Match 92.8%; Score 1879.6; DB 13; Length 2177;	
XX	Best Local Similarity 96.2%; Pred. No. 0;	
XX	Matches 1946; Conservative 0; Mismatches 74; Indels 2; Gaps 2;	
XX		
XX	4 GGAGCTTGAGGAAACGAGAGGAAAGAGAGAAATGGCGTCCAGAGATTACAGTACCTAT 63	
XX	1 GGAGCTTGAGGAAACGAGAGGAAAGAGAGAAATGGCGTCCAGAGATTACAGTACCTAT 60	
XX	64 AGCCAGGTGAGGAGGAGCAGGAGGCTACAGTCTTACACCGCCAGGCCACTCAAGATAT 123	
XX	61 AGCCAGGTGAGGAGGAGCAGGAGGCTACAGTCTTACACCGCCAGGCCACTCAAGATAT 120	
XX	124 GCACGAGCACCCAGGCATATGCGCAACAAAGCTATGGAACCTATGACAGCCCACTGAT 183	
XX	121 GCACGAGCACCCAGGCATATGCGCAACAAAGCTATGGAACCTATGACAGCCCACTGAT 180	
XX	184 GTCAAGTATACCCAGAGCTCAGACCACTGATGGCGAGACCGCTTATGCAACTTCT 243	
XX	181 GTCAAGTATACCCAGAGCTCAGACCACTGATGGCGAGACCGCTTATGCAACTTCT 240	
XX	244 TATGAGACGCTCCCACTGGTTATATCTCAACTGCCCCCAGGGCATACAGCGAGCT 303	
XX	241 TATGAGACGCTCCCACTGGTTATATCTCAACTGCCCCCAGGGCATACAGCGAGCT 300	
XX	304 GTCCAGGGGTATGGAATGCGATGCTTATATACCACTGCTACAGTACCAACCAACCGAG 363	
XX	301 GTCCAGGGGTATGGAATGCGATGCTTATATACCACTGCTACAGTACCAACCAACCGAG 360	
XX	364 GCGTCTTATGCACTCACTGCTGATATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 423	
XX	361 GCGTCTTATGCACTCACTGCTGATATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 420	
XX	424 CAGCGAGACGACCTGACCTTACCAAGACCGGAGATGGAACCAAGCCCACTGAGACTAGT 483	

421 GAGCCAGTACGCTGACCTTACAGCTGACGATGAAACAGCCCACTGAGACTAGT 480  
484 CAACCTCAATCTAGCAGAGGGGTTTACAAACAGCCCACTGAGATATGACAGAGTAC 543  
481 CAACCTCAATCTAGCAGAGGGGTTTACAAACAGCCCACTGAGATATGACAGAGTAC 540  
544 TACAGTTATCCCAAGGTAAGTGGAGGTAACCCATGACAGGCACTGACCTCCATCC 603  
541 TGAAGTTATCCCAAGGTAAGTGGAGGTAACCCATGACAGGCACTGACCTCCATCC 600  
604 TACCTCTCAAGCTATCTCTTACAGAGGCACTGATATGATGACAGAGTACTCT 663  
601 TACCTCTCAAGCTATCTCTTACAGAGGCACTGATATGATGACAGAGTACTCT 660  
664 CAGCAGAACCACTATGAGGCAACCGAGAGCTATGACAGAGTATGATGATGATCA 723  
661 CAGCAGAACCACTATGAGGCAACCGAGAGCTATGACAGAGTATGATGATGATCA 720  
724 AGCAGCTATGAGGCAAGGCTCCCACTATGATGATGATGATGATGATGATGATG 783  
721 AGCAGCTATGAGGCAAGGCTCCCACTATGATGATGATGATGATGATGATGATG 780  
784 GCTCAAGTCAATATAGCAACAGAGAGCTAGAGGAGAGGAGTTCATCCAGCAG 843  
781 GCTCAAGTCAATATAGCAACAGAGAGCTAGAGGAGAGGAGTTCATCCAGCAG 840  
844 GACCACTCCAGTACATGAGGAGGTTTATGAGGAGAGGAGTTCATCCAGCAG 903  
841 GACCACTCCAGTACATGAGGAGGTTTATGAGGAGAGGAGTTCATCCAGCAG 900  
904 GAGAACCGAGAGCATGAGTGGGCTGATTAACCGGAGAGGAGGAGGAGGAGTTCAT 963  
901 GAGAACCGAGAGCATGAGTGGGCTGATTAACCGGAGAGGAGGAGGAGGAGTTCAT 960  
964 GAGAGCATGAGAGAGGTTGGGCGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 1023  
961 GAGAGCATGAGAGAGGTTGGGCGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 1020  
1024 CGAGGTGGCTTCAATAGCTGAGTGAACCAATGATGAGAGGAGGAGGAGGAGG 1083  
1021 CAAGTTGGCTTCAATAGCTGAGTGAACCAATGATGAGAGGAGGAGGAGGAGG 1080  
1084 CCACCTGTATATCAGATGAGAGGAGTCTGACAACTGATGATGATGATGATGAT 1143  
1081 CCACCTGTATATCAGATGAGAGGAGTCTGACAACTGATGATGATGATGATGAT 1140  
1144 GACAGTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1203  
1141 GACAGTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
1204 AACAGAGAACTGGGCAACCAATGATCAGATCTGATGATGATGATGATGATGATG 1263  
1201 AACAGAGAACTGGGCAACCAATGATCAGATCTGATGATGATGATGATGATGATG 1260  
1264 AAAGGCGATGCCAGATGCTCTGATGAGAGGAGGAGGAGGAGGAGGAGGAGG 1323  
1261 AAAGGCGATGCCAGATGCTCTGATGAGAGGAGGAGGAGGAGGAGGAGGAGG 1320  
1324 TTGATGAGGAGAAATTTTCAAGGAGAGCAATTAATCTCTGATGATGATGATGATG 1383  
1321 TTGATGAGGAGAAATTTTCAAGGAGAGCAATTAATCTCTGATGATGATGATGATG 1380  
1384 CCAATGAAAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1443  
1381 CCAATGAAAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
1444 CTCTGCTGAGAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1503  
1441 CTCTGCTGAGAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
1504 CGTGAAGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1563

1501 CGTGAAGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
1564 TCTGAGAGAGAGAAAGCTGACGACCGAGCTGAGAGCTGAGAGTGTCCAAATCCGGTGT 1623  
1561 TCTGAGAGAGAGAAAGCTGACGACCGAGCTGAGAGCTGAGAGTGTCCAAATCCGGTGT 1620  
1624 GGAACCAAGAACTTCCGCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1683  
1621 GGAACCAAGAACTTCCGCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
1684 GGCCTCTCCGCGACACCTTCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1743  
1681 GGCCTCTCCGCGACACCTTCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740  
1744 GGCATGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1803  
1741 GGCATGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800  
1804 GGTGGCTGT 1863  
1801 GGTGGCTGT 1859  
1864 TTTGT 1923  
1860 TTTGT 1918  
1924 GGAAG 1983  
1919 GGAAG 1978  
1984 GAGCGAGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2025  
1979 GAGCGAGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2020

RESULT 11  
AAS62623/c  
ID AAS62623 standard; cDNA; 2273 BP.  
XX  
AC AAS62623;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
DE cDNA sequence #410 encoding novel human secreted protein.  
XX  
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
KW immune deficiency disorder; blood disorder; inflammatory disorder;  
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
KW immunosuppressive; antineumatic; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200177291-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 29-MAR-2001; 2001WO-US010485.  
XX  
PR 06-APR-2000; 2000US-0195604P.  
XX  
PA (GEMV) GENETICS INST INC.  
XX  
PI Wong GG, Clark HF, Recheil K, Agostino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham JR;  
XX  
DR WPI; 2002-010900/01.  
XX  
XX New polynucleotides encoding secreted proteins useful for treating e.g.  
PT asthma, HIV and Crohn's disease.  
XX  
PS Claim 1; Page 291; 391pp; English.  
XX  
CC The present invention relates to the isolation of novel cDNA sequences

CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. AAS62214-AAS62838  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins  
CC  
SQ Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;  
Query Match 91.7%; Score 1856.6; DB 6; Length 2273;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1933; Conservative 0; Mismatches 4; Indels 54; Gaps 2;  
QY 35 AAATGGGCTCAGCGATTACAGTACTTATAGCCAGCTGACGCGACAGAGGGCTACAGTG 94  
DB 2268 AAGGTGGCTCAGCGATTACAGTACTTATAGCCAGCTGACGCGACAGAGGGCTACAGTG 2209  
QY 95 CTTACACCGCCGACGCTCAGAGATATGACAGACACCCAGGCTATGGGCAACAA 154  
DB 2208 CTTACACCGCCGACGCTCAGAGATATGACAGACACCCAGGCTATGGGCAACAA 2149  
QY 155 GCTATGGAACCTATGACAGAGCCAGCTAGTCACTATACCCAGGCTCAGACCACTGCA 214  
DB 2148 GCTATGGAACCTATGACAGAGCCAGCTAGTCACTATACCCAGGCTCAGACCACTGCA 2089  
QY 215 CCTATGGGACAGCCGCTATGCACTTCTATGAGACGCTCCCACTGGTTATCTACTC 274  
DB 2088 CCTATGGGACAGCCGCTATGCACTTCTATGAGACGCTCCCACTGGTTATCTACTC 2029  
QY 275 CAACTGCCCCCGAGGCTATACAGCGCTGTCAGGGGTATGGACTGGTCTTATATATA 334  
DB 2028 CAACTGCCCCCGAGGCTATACAGCGCTGTCAGGGGTATGGACTGGTCTTATATATA 1969  
QY 335 CCAACACTGCTACAGTACCAACCAACAGGCGCTCTATGACGCTCACTGTCATATGCA 394  
DB 1968 CCAACACTGCTACAGTACCAACCAACAGGCGCTCTATGACGCTCACTGTCATATGCA 1909  
QY 395 CTCAGCGCTGCTTATCCAGGCTATGGGACAGGACGACCTGCACTTACAAAGCCG 454  
DB 1908 CTCAGCGCTGCTTATCCAGGCTATGGGACAGGACGACCTGCACTTACAAAGCCG 1849  
QY 455 AGGATGGAACCAAGCCCACTGAGACTAGTCACTTCAATCTAGCAAGGGGGTTTACAA 514  
DB 1848 AGGATGGAACCAAGCCCACTGAGACTAGTCACTTCAATCTAGCAAGGGGGTTTACAA 1789  
QY 515 AGCCAGCGCTATGATGACAGAGTAACTACAGTTATCCCAAGTACTGGAGTACC 574  
DB 1788 AGCCAGCGCTATGATGACAGAGTAACTACAGTTATCCCAAGTACTGGAGTACC 1729  
QY 575 CCATGACGACGATGACCTTCATCTTCACTCTCTTCAAGGTTATCTTACAGAGC 634  
DB 1728 CCATGACGACGATGACCTTCATCTTCACTCTCTTCAAGGTTATCTTACAGAGC 1669  
QY 635 CGACTATGATGATCAAGGCTTAACTTCAAGCAACCACTATGGGCAACCGAGAGCT 694  
DB 1668 CGACTATGATGATCAAGGCTTAACTTCAAGCAACCACTATGGGCAACCGAGAGCT 1609  
QY 695 ATGACAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 754  
DB 1608 ATGACAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549  
QY 755 ACCCAACCCCAAACTGATCTTACAGCCAAAGCTTCAAGTCAATATATAGCCAAAGAGCA 814  
DB 1548 ACCCAACCCCAAACTGATCTTACAGCCAAAGCTTCAAGTCAATATATAGCCAAAGAGCA 1489

QY 815 GCTACGGGAGAGAGAGTATCCGACAGAGCAACCCAGTACATGGGTTTATAGGC 874  
DB 1488 GCTACGGGAGAGAGAGTATCCGACAGAGCAACCCAGTACATGGGTTTATAGGC 1429  
QY 875 AGGAGCTCTGAGAGATTTTCCGACAGAGAGAACCGAGCATAGTGGCTTGATPACC 934  
DB 1428 AGGAGCTCTGAGAGATTTTCCGACAGAGAGAACCGAGCATAGTGGCTTGATPACC 1369  
QY 935 GGGGACGGGGAAGAGGGGATTTGATCTGAGAGGCTATGACAGAGTGGCGGGGAGAG 994  
DB 1368 GGGGACGGGGAAGAGGGGATTTGATCTGAGAGGCTATGACAGAGTGGCGGGGAGAG 1309  
QY 995 GACGGGATGAAATGGGACGGCTGAGAGAGAGTGGCTTCAATTAAGCTGTGACCCA 1054  
DB 1308 GACGGGATGAAATGGGACGGCTGAGAGAGAGTGGCTTCAATTAAGCTGTGACCCA 1252  
QY 1055 TGGATGAGAGACAGATCTTGAATCTAGGCCCACCTTGAATCCAGATGAAGCTTGACA 1114  
DB 1251 TGGATGAGAGACAGATCTTGAATCTAGGCCCACCTTGAATCCAGATGAAGCTTGACA 1192  
QY 1115 ACAGTCAATTTATGTAAGAGATTAATGACAGTGTGACTTGAATGATCTGGCAGACT 1174  
DB 1191 ACAGTCAATTTATGTAAGAGATTAATGACAGTGTGACTTGAATGATCTGGCAGACT 1132  
QY 1175 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAAGAACTGGGCAACCATGATCCACA 1234  
DB 1131 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAAGAACTGGGCAACCATGATCCACA 1072  
QY 1235 TCTTACCTGACCAAGAAACAGAAAGCCAAAGGCGATGACAGTGTCTTATGAAGACC 1294  
DB 1071 TCTTACCTGACCAAGAAACAGAAAGCCAAAGGCGATGACAGTGTCTTATGAAGACC 1012  
QY 1295 CACCCACTGCGAAGGCTGCGGTGGAATGTTTGAATGGAAAGTTTCAAGGGAAGCAAC 1354  
DB 1011 CACCCACTGCGAAGGCTGCGGTGGAATGTTTGAATGGAAAGTTTCAAGGGAAGCAAC 952  
QY 1355 TTAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATCGGGGTGTCTGAC 1414  
DB 951 TTAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATCGGGGTGTCTGAC 892  
QY 1415 CCGGTGAGGAGAGGACATGACCAACCACTCCGTGAGGTCCAGAGGCCAGAGAGTCC 1474  
DB 891 CCGGTGAGGAGAGGACATGACCAACCACTCCGTGAGGTCCAGAGGCCAGAGAGTCC 832  
QY 1475 CTGGGGAACCCATGGGTTCGATGGAGGCGGTGGAAGATGAAGAGGCTTCCCTCA 1534  
DB 831 CTGGGGAACCCATGGGTTCGATGGAGGCGGTGGAAGATGAAGAGGCTTCCCTCA 772  
QY 1535 GAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAAAGTTCACACCGAGCTG 1594  
DB 771 GAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAAAGTTCACACCGAGCTG 712  
QY 1595 GAGACTGGCAGTGTCCCAATCCGGGTTTGAAGAACCAAACTTCCCTGAGAAACAGAGT 1654  
DB 711 GAGACTGGCAGTGTCCCAATCCGGGTTTGAAGAACCAAACTTCCCTGAGAAACAGAGT 652  
QY 1655 GCAACCAAGTAAAGGCCCAAAAGCTGAGAGCTTCTCCGCAACCTTCCGCCCCGG 1714  
DB 651 GCAACCAAGT----- 643  
QY 1715 GTGGTATCGTGGCAGAGTGGCCCTGTGGGCAATGCGGGAGAGAAAGAGTGGCTTATGG 1774  
DB 642 GTGGTATCGTGGCAGAGTGGCCCTGTGGGCAATGCGGGAGAGAAAGAGTGGCTTATGG 583  
QY 1775 ATCGTGTGTGTCCTCGGTGAATGTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCC 1834  
DB 582 ATCGTGTGTGTCCTCGGTGAATGTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCC 523  
QY 1835 GTGGTGGCCGGGCTATGACCCAGAGTGGCTTGTGGAGAGAAACAGAGTGGCCCTGGGG 1894  
DB 522 GTGGTGGCCGGGCTATGACCCAGAGTGGCTTGTGGAGAGAAACAGAGTGGCCCTGGGG 463







```
QY 506 GTTAAACACGACCCAGCTAGATATGAGACAGAGTAACTACAGTTATCCCGAGTACTG 565
Db 423 ----- 422
QY 566 GGAAGTACCCCATGACAGCAGTCACTGCACTCCATCTCACTCACTCACTATTTCCT 625
Db 423 -----AGCTATTTCCT 432
QY 626 CTACACAGCCGACTAGTTATGATCAGACAGTAACTCTCAGCAGAAACAAGCTATGGGCAAC 685
Db 433 CTACACAGCCGACTAGTTATGATCAGACAGTAACTCTCAGCAGAAACAAGCTATGGGCAAC 492
QY 686 CGAGCAGCTATGAGCAGCAGAGTACGTTAGTGTCAACAAAGAGCTATGGGAGCAGCCTC 745
Db 493 CGAGCAGCTATGAGCAGCAGAGTACGTTAGTGTCAACAAAGAGCTATGGGAGCAGCCTC 552
QY 746 CCACTAGTTAACCAACCCCAAACTGATCTCTCAGCAGCAAGCTCCAAGTCAATATATAGCCAAC 805
Db 553 CCACTAGTTAACCAACCCCAAACTGATCTCTCAGCAGCAAGCTCCAAGTCAATATATAGCCAAC 612
QY 806 AGAGCAGCAGCTACGAGGAGCAGAGATTCAATCCGACAGAACCAACCCAGTAGACAGGAGT 865
Db 613 AGAGCAGCAGCTACGAGGAGCAGAGATTCAATCCGACAGAACCAACCCAGTAGACAGGAGT 672
QY 866 TTTATGGGAGCAGAGTCTGAGAGATTTTCCGACCAAGAGAGAACCCGAGCATGAGTGGCC 925
Db 673 TTTATGGGAGCAGAGTCTGAGAGATTTTCCGACCAAGAGAGAACCCGAGCATGAGTGGCC 732
QY 926 CTGATTAACCGGGAGCAGGGAGAAAGAGGGGAAATTGATCGTGGAGGATGAGCAGAGTGGGC 985
Db 733 CTGATTAACCGGGAGCAGGGAGAAAGAGGGGAAATTGATCGTGGAGGATGAGCAGAGTGGGC 792
QY 986 GGGAGAGAGAGACCGCGGTGAAATGGGCAAGCGCTGAGAGAGCAGAGTGGCTTCAATTAAGCTG 1045
Db 793 GGGAGAGAGAGACCGCGGTGAAATGGGCAAGCGCTGAGAGAGCAGAGTGGCTTCAATTAAGCTG 852
QY 1046 GTGAGACCATGAGTAAAGAGACCAATCTTGATCTAGAGCCCACTGTGATGTCAGATGAG 1105
Db 853 GTGAGACCATGAGTAAAGAGACCAATCTTGATCTAGAGCCCACTGTGATGTCAGATGAG 912
QY 1106 ACTGTGACAAAGTGAATTTATGATACAGAGATTTAAATGACAGTGTGCTTAATGATGTC 1165
Db 913 ACTGTGACAAAGTGAATTTATGATACAGAGATTTAAATGACAGTGTGCTTAATGATGTC 972
QY 1166 TGGCAGACTCTTTAAGCAGTGTGGGTGTTAAGATGAAACAGAGAACTGGGCAACCA 1225
Db 973 TGGCAGACTCTTTAAGCAGTGTGGGTGTTAAGATGAAACAGAGAACTGGGCAACCA 1032
QY 1226 TGATCCACATCTACCTGAGCAAGAAACAGAAAGCCCAAGGATGCCACAGTGTCT 1285
Db 1033 TGATCCACATCTACCTGAGCAAGAAACAGAAAGCCCAAGGATGCCACAGTGTCT 1092
QY 1286 ATGAAAGACCAACCACTGACCAAGGCTCGGAGAAAGTGTGATGGGAAAGATTTTCAAG 1345
Db 1093 ATGAAAGACCAACCACTGACCAAGGCTCGGAGAAAGTGTGATGGGAAAGATTTTCAAG 1152
QY 1346 GGAAGCAAACTTAAAGTCTCCCTTGTGCGAAGAAAGCTCCATGATGAAGATGCGGGGTG 1405
Db 1153 GGAAGCAAACTTAAAGTCTCCCTTGTGCGAAGAAAGCTCCATGATGAAGATGCGGGGTG 1212
QY 1406 GTCTGCCACCCCGTGAAGGAGCAGAGCATGCCACCACTCTCGTGAAGTCCAGAGAGCC 1465
Db 1213 GTCTGCCACCCCGTGAAGGAGCAGAGCATGCCACCACTCTCGTGAAGTCCAGAGAGCC 1272
QY 1466 CAGAGAGTCTGGGGAGACCAATGGGTGCAATGGAGGCGGAGGAGATGAGAGAGGCT 1525
Db 1273 CAGAGAGTCTGGGGAGACCAATGGGTGCAATGGAGGCGGAGGAGATGAGAGAGGCT 1332
QY 1526 TCCCTCCAAAGAGACCCCGGGGTTCGCCAGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGC 1585
Db 1333 TCCCTCCAAAGAGACCCCGGGGTTCGCCAGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGC 1392
```

```
QY 1586 ACCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGTTGTGAAACAGAACTTCCCTGGA 1645
Db 1393 ACCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGTTGTGAAACAGAACTTCCCTGGA 1452
QY 1646 GAAACAGAGTCAACCAAGTGAAGGCCCCAAAGCCTGAAGGCTTCTCCCGCACCTTTTC 1705
Db 1453 GAAACAGAGTCAACCAAGTGAAGGCCCCAAAGCCTGAAGGCTTCTCCCGCACCTTTTC 1512
QY 1706 CGCCCTCGGGTGTGATCTGGGAGAGAGTGGCCCTGTGGCATGCGGGAGAAAGAGGTG 1765
Db 1513 CGCCCTCGGGTGTGATCTGGGAGAGAGTGGCCCTGTGGCATGCGGGAGAAAGAGGTG 1572
QY 1766 GCTCATGAGATCCGATGATGATCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAG 1825
Db 1573 GCTCATGAGATCCGATGATGATCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAG 1632
QY 1826 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1885
Db 1633 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1692
QY 1886 GCCCTGGGGGGCCCCCTGACCTTTGATGAAACAGATGGGAGAAAGAGAGAGAGCTG 1945
Db 1693 GCCCTGGGGGGCCCCCTGACCTTTGATGAAACAGATGGGAGAAAGAGAGAGAGCTG 1752
QY 1946 GAGGACCTGGAATAAATGATTAAGGCGAGCAGCCTGACGAGAGCGAGATCGGCCCTACT 2005
Db 1753 GAGGACCTGGAATAAATGATTAAGGCGAGCAGCCTGACGAGAGCGAGATCGGCCCTACT 1812
QY 2006 AGATGCAAGAGACCCCGCAGA 2025
Db 1813 AGATGCAAGAGACCCCGCAGA 1832

RESULT 14
ADRO7446
ID ADRO7446 standard; cDNA; 2026 BP.
XX
AC ADRO7446;
XX
DT 04-NOV-2004 (first entry)
XX
DE Full length human cDNA useful for treating neurological disease Seq 952.
XX
KW gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
KW tranquiliser.
XX
OS Homo sapiens.
XX
PN BP1447413-A2.
XX
PD 18-AUG-2004.
XX
PF 12-FEB-2004; 2004BP-00003145.
XX
PR 14-FEB-2003; 2003JP-00102207.
XX
PR 09-MAY-2003; 2003JP-00131452.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX
DR WPI; 2004-583265/57.
XX
DR P-PSDB; ADRO9402.
XX
XX
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
XX Claim 1; SEQ ID NO 952; 2686bp; English.
```



Oy	1801	GGAGGTGGCCCGTGGTGGAGACAAGAGGTGGCTTCGCTGGTGGCCCGGGGCGATGACCGAGGT	1860
Db	1646	AGAGGTGCCCTGTGGTGGACAAGAGGTGGCTTCGCTGGTGGCCCGGGGCGATGACCGAGGT	1705
Oy	1861	GGCTTTGGTGGAGCAGAGCGAGGTGGCCCTCGGGGGGCCCTCGACCTTTGATGGACAG	1920
Db	1706	GGCTTTGGTGGAGGAAAGACGAGGTGGCCCTGGGGGGGCCCTCGACCTTTATGGACAG	1765
Oy	1921	ATGGGAGGAAAGAGAGGAGACGTGGAGGACCTGGAAAAATGATTAAGGCGAGCACCGT	1980
Db	1766	ATGGGAGGAAAGAGAGGAGACGTGGAGGACCTGGAAAAATGATTAAGGCGAGCACCGT	1825
Oy	1981	CAGGAGCGCAGAGATCGGCCCTACATGATGGACAGAGCCCGCAG	2025
Db	1826	CAGGAGCGCAGAGATCGGCCCTACATGATGGACAGAGCCCGCAG	1870

RESULT 15  
 ADP56333  
 ID ADP56333 standard; cDNA; 1807 BP.  
 XX  
 AC ADP56333;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human PRO cDNA sequence SEQ ID NO:2309.  
 XX  
 KM human; PRO; immune related disease; inflammatory immune response;  
 KM immune response stimulation; antiallergic; antianaemic; antiarthritic;  
 KM antidiabetic; antidiabetic; antiinflammatory; antipsostatic;  
 KM antirheumatic; antihypoid; CNS; dermatological; gastrointestinal;  
 KM haemotropic; hepatotropic; immunostimulant; immunosuppressive; muscular;  
 KM neurotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
 KM vitruide; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004039956-A2.  
 XX  
 PD 13-MAY-2004.  
 XX  
 PF 28-OCT-2003; 2003WO-US034381.  
 XX  
 PR 29-OCT-2002; 2002US-0422472P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;  
 PI Wood WI, Wu TD;  
 XX  
 DR WPI; 2004-376182/35.  
 DR P-PSDB; ADP56334.  
 PT  
 PT New PRO polynucleotides and polypeptides, useful in useful in diagnosing  
 PT and treating an immune related disease, e.g. systemic lupus  
 PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in  
 PT stimulating an immune response.  
 XX  
 Claim 2; SEQ ID NO 2309; 3009bp; English.  
 XX  
 XX The present invention describes an isolated PRO nucleic acid (I). Also  
 CC described: (1) a vector comprising (II) (2) a host cell comprising the  
 CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an  
 CC isolated PRO polypeptide; (5) a chimeric molecule comprising the  
 CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
 CC antibody which specifically binds to a polypeptide of (4); (7) a  
 CC composition of matter comprising a polypeptide of (4); an agonist or  
 CC antagonist of the polypeptide or an antibody that binds to the  
 CC polypeptide in combination with a carrier; (8) an article of manufacture  
 CC comprising a container, a label on the container and a composition of  
 CC matter of (7); (9) a method of treating an immune related disease in a  
 CC mammal; (10) a method for determining the presence of a PRO polypeptide  
 CC in a sample suspected of having the polypeptide; (11) a method of

CC diagnosing an immune related disease or an inflammatory immune response  
CC in mammal; (12) a method of identifying a compound that inhibits or  
CC mimics the activity of or expression of a gene encoding a PRO polypeptide  
CC and (13) a method of stimulating the immune response in a mammal. The  
CC PRO sequences have antiallergic, antianemic, antarthritic,  
CC antiasthmatic, antidiabetic, antiinflammatory, antipsoriatic,  
CC antirheumatic, antithyroid, CNS, dermatological, gastrointestinal,  
CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,  
CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and  
CC vitreous activities, and can be used in gene therapy. The nucleic acid  
CC (I) and the encoded polypeptides, compositions, kits and methods are  
CC useful in diagnosing and treating an immune related disease and in  
CC stimulating an immune response. The present sequence represents a human  
CC PRO nucleotide sequence from the present invention.  
XX  
SQ Sequence 1807 BP, 476 A, 497 C, 523 G, 311 T, 0 U, 0 Other;

Query Match	76.9%;	Score 1557.4;	DB 13;	Length 1807;
Best Local Similarity	89.0%;	Pred. No. 0;		
Matches 1787;	Conservative	0;	Mismatches 1;	Indels 219;
			Gaps	1;

OY	1	GACGACGCTTAGAAGAACAGAGAGAAAGAGAAATAATGGGCTCCAGATTAACGTAC	60
Db	20	GACGACGCTTAGAGAGAACAGAGAGAGAGAGAGAAAATGGGCTCCAGATTAACGTAC	79
OY	61	TATAGCCAAAGCTGACGCGACAGACGAGGCTTACAGTGTCTTACACCGGCCAGCCCTCAAGGA	120
Db	80	TATAGCCAAAGCTGACGCGACAGACGAGGCTTACAGTGTCTTACACCGGCCAGCCCTCAAGGA	139
OY	121	TATGACACAGACACCCACGACGCTATATGGGCAACAAAGCTATGGAACTTATGGACAGCCACT	180
Db	140	TATGACACAGACACCCACGACGCTATATGGGCAACAAAGCTATGGAACTTATGGACAGCCACT	199
OY	181	GATGTCAGACTATACCCAGGCTCAGACCACTGCAACTATGGGCGAGACGCTTATGCAACT	240
Db	200	GATGTCAGACTATACCCAGGCTCAGACCACTGCAACTATGGGCGAGACGCTTATGCAACT	259
OY	241	TCTTATGGACAGCCTCCCACTGGTTATATCTACCTCCAACTGCCCCCAGGACATACGCGAG	300
Db	260	TCTTATGGACAGCCTCCCACTGGTTATATCTACCTCCAACTGCCCCCAGGACATACGCGAG	319
OY	301	CCTGTCCAGGGGGTATGGGCACTGGTGTCTTATGATACACCACTGCTACAGTACACCAACC	360
Db	320	CCTGTCCAGGGGGTATGGGCACTGGTGTCTTATGATACACCACTGCTACAGTACACCAACC	379
OY	361	CAGGCGCTCTATGAGGCTCAGTCTGCAATATGGGCACTAGCCTGTCTTATCCAGCCTATGGG	420
Db	380	CAGGCGCTCTATGAGGCTCAGTCTGCAATATGGGCACTAGCCTGTCTTATCCAGCCTATGGG	439
OY	421	CAGGACGACAGGACCACTGCACTTACACAAAGCCGAGGATGGAAACAAGCCCACTGAGACT	480
Db	440	CAGGACGACAGGACCACTGCACTTACACAAAGCCGAGGATGGAAACAAGCCCACTGAGACT	499
OY	481	AGTCAACCTCAATTTAGACAAGGGGGGTTACAACGAGCCAGCCTTAGATATGACAGAGT	540
Db	500	AGTCAACCTCAATTTAGACAAGGGGGGTTACAACGAGCCAGCCTTAGATATGACAGAGT	559
OY	541	AACATCAAGTTATCCCGAGGTACCTGGGAGGCTACCCCATGGACGACCACTGCACTGCCA	600
Db	560	AACATCAAGTTATCCCGAGGTACCTGGGAGGCTACCCCATGGACGACCACTGCACTGCCA	619
OY	601	TCTTACCTCCTTACAGACTATTCCTCTTACACAGCGGACTAGTTATGATACAGAGCAGTTAC	660
Db	620	TCTTACCTCCTTACAGACTATTCCTCTTACACAGCGGACTAGTTATGATACAGAGCAGTTAC	679
OY	661	TCTTACAGAAACACCTTATGGGCAACCGAGCACTATGGAACGACAGAGTACTATGTCTAAC	720
Db	680	TCTTACAGAAACACCTTATGGGCAACCGAGCACTATGGAACGAGAGTACTATGTCTAAC	739
OY	721	CAAAGCAGCTATGGGCGACGCGCTCCCACTAGTTACCAACCCAAACGTGATCTCTAACG	780
Db	740	CAAAGCAGCTATGGGCGACGCGCTCCCACTAGTTACCAACCCAAACGTGATCTCTAACG	799

QY 781 CAAGCTCCAGTCAATATAGCAACAGAGGAGAGCTACGGGAGAGAGATTCAATCCGA 840  
DB 800 CAAGCTCCAGTCAATATAGCAACAGAGGAGAGCTACGGGAGAGAGATTCAATCCGA 848  
QY 841 CAGAGACCAACCCAGTACATGAGTGTATATGAGGAGAGAGTCTGAGAGATTTCCGAGCA 900  
DB 849 ----- 848  
QY 901 GAGAGAACCGGAGAGATGAGTGGCTTGAATACCGGGGAGAGGAGAGATTTGAT 960  
DB 849 ----- 848  
QY 961 CGTGAAGAGATGAGCAGAGAGTGGCGGGAGAGAGACCGGAGATGAGAGAGCTGGA 1020  
DB 849 ----- 848  
QY 1021 GAGGAGGTGGCTTCAATTAAGCTGTGAGACCCATGATGAGAGAGAGATCTTGATCTA 1080  
DB 849 -----GACCATGATGAGAGACAGATCTTGATCTA 880  
QY 1081 GGGCCACCTGTAGATCCAGATGAAGATCTGACACAGTGCATTTATGTCAGAGATTA 1140  
DB 881 GGGCCCTCTGTAGATCCAGATGAAGATCTGACACAGTGCATTTATGTCAGAGATTA 940  
QY 1141 AATGACAGTGTGATCTAGATGATCTGAGAGATCTTCTTAAAGCAGTGTGGGAGTTGTTAAG 1200  
DB 941 AATGACAGTGTGATCTAGATGATCTGAGAGATCTTCTTAAAGCAGTGTGGGAGTTGTTAAG 1000  
QY 1201 ATGAAACAAGAACTGAGGCAACCCATGATCCATCTACCTGAGCAAGAAACAGAAAG 1260  
DB 1001 ATGAAACAAGAACTGAGGCAACCCATGATCCATCTACCTGAGCAAGAAACAGAAAG 1060  
QY 1261 CCCAAAGGCGATGCAAGATGTCTATGAAGACCCCACTGCAAGAGCTGCGGTGGA 1320  
DB 1061 CCCAAAGGCGATGCAAGATGTCTATGAAGACCCCACTGCAAGAGCTGCGGTGGA 1120  
QY 1321 TGGTTGATGGGAAAGATTTTCAAGGAGAGCAACTTAAAGTCTCTTGTCTCGAAGAG 1380  
DB 1121 TGGTTGATGGGAAAGATTTTCAAGGAGAGCAACTTAAAGTCTCTTGTCTCGAAGAG 1180  
QY 1381 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAGAGGAGAGCATGCCA 1440  
DB 1181 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAGAGGAGAGCATGCCA 1240  
QY 1441 CCACTCCGTGAGAGTCCAGAGGAGCCAGAGAGTCTTGGGAGACCCATGGGTCCGATGGGA 1500  
DB 1241 CCACTCCGTGAGAGTCCAGAGGAGCCAGAGAGTCTTGGGAGACCCATGGGTCCGATGGGA 1300  
QY 1501 GGCCTGAGAGAGATGAGAGAGGCTTCTCTCAAGAGAGACCCCGGGGTCCCGAGGGAGAC 1560  
DB 1301 GGCCTGAGAGAGATGAGAGAGGCTTCTCTCAAGAGAGACCCCGGGGTCCCGAGGGAGAC 1360  
QY 1561 CCTCTGAGAGAGAGAGAGTCCAGACAGAGCTGAGAGCTGAGAGTCTCCCATCCGGGT 1620  
DB 1361 CCTCTGAGAGAGAGAGAGTCCAGACAGAGCTGAGAGCTGAGAGTCTCCCATCCGGGT 1420  
QY 1621 TGTGAGAAACCAAGACTTCCCTGAGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGACT 1680  
DB 1421 TGTGAGAAACCAAGACTTCCCTGAGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGACT 1480  
QY 1681 GAAGGCTTCTCCCGGACCTTCCGCGCCCGGGGTGTGATCGTGGCAGAGGTGGCCT 1740  
DB 1481 GAAGGCTTCTCCCGGACCTTCCGCGCCCGGGGTGTGATCGTGGCAGAGGTGGCCT 1540  
QY 1741 GGTGGCATGCGGAGAGAGAGTGGCTCATGATCGTGTGTCCCGGTGAGATGTTT 1800  
DB 1541 GGTGGCATGCGGAGAGAGAGTGGCTCATGATCGTGTGTCCCGGTGAGATGTTT 1600  
QY 1801 AGAGGTGGCTGTGTGAGAGACAGAGTGGCTTCCGTGTGGCCGGGAGATGAGACCGAGGT 1860  
DB 1601 AGAGGTGGCTGTGTGAGAGACAGAGTGGCTTCCGTGTGGCCGGGAGATGAGACCGAGGT 1660  
QY 1861 GGCCTTGTGTGAGAGAGACAGAGTGGCTGTGGGGGCCCCCTGAGACCTTTGATGAGACAG 1920

DB 1661 GGCCTTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGACCTTTGATGAGACAG 1720  
QY 1921 ATGGAGAGAGAGAGAGAGAGTGGAGACTTGAAGAAATGATTAAGGCGAGACCGT 1980  
DB 1721 ATGGAGAGAGAGAGAGAGAGTGGAGACTTGAAGAAATGATTAAGGCGAGACCGT 1780  
QY 1981 CAGAGCGCAGAGATGAGGCTTACTAG 2007  
DB 1781 CAGAGCGCAGAGATGAGGCTTACTAG 1807

Search completed: February 20, 2005, 05:30:45  
Job time : 997.09 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53 ; Search time 335.821 Seconds  
(without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_8\_2032

Perfect score: 2025  
Sequence: 1 gacgcagcgtgagagacga.....agatgcagagacccgcagaga 2025

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81818359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2011.4	99.3	2371	2 US-08-343-443B-1	Sequence 1, Appl1
2	1884.4	93.1	6002	4 US-09-949-016-13696	Sequence 13696, A
3	1556	76.8	1785	4 US-09-949-016-5043	Sequence 5043, Ap
4	1556	76.8	1785	4 US-09-949-016-5044	Sequence 5044, Ap
5	1429.6	70.6	1783	4 US-09-949-016-1954	Sequence 1954, Ap
6	819	40.4	2412	1 US-08-437-027-18	Sequence 18, Appl
7	424	20.9	454	4 US-09-513-999C-1657	Sequence 1657, Ap
8	364	18.0	601	4 US-09-949-016-66382	Sequence 66382, A
9	338	16.7	411	4 US-09-621-876-13361	Sequence 13361, A
10	299.2	14.8	393	4 US-09-513-999C-12062	Sequence 12062, A
11	266.4	13.2	954	2 US-08-343-443B-106	Sequence 106, App
12	258.4	12.8	35784	4 US-09-949-016-16785	Sequence 16785, A
13	258.4	12.8	35784	4 US-09-949-016-16786	Sequence 16786, A
14	213	10.5	601	4 US-09-949-016-176641	Sequence 176641, A
15	213	10.5	601	4 US-09-949-016-176644	Sequence 176644, A
16	175.2	8.7	1939	4 US-09-919-039-322	Sequence 322, App
17	135.8	6.7	601	4 US-09-949-016-176637	Sequence 176637, A
18	135.8	6.7	601	4 US-09-949-016-176638	Sequence 176638, A
19	135.8	6.7	601	4 US-09-949-016-176660	Sequence 176660, A
20	135.8	6.7	601	4 US-09-949-016-176661	Sequence 176661, A
21	129.4	6.4	601	4 US-09-949-016-176622	Sequence 176622, A
22	129.4	6.4	601	4 US-09-949-016-176623	Sequence 176623, A
23	129.4	6.4	601	4 US-09-949-016-176645	Sequence 176645, A
24	129.4	6.4	601	4 US-09-949-016-176646	Sequence 176646, A
25	116.4	5.7	450	4 US-09-370-838-145	Sequence 145, App
26	116.4	5.7	450	4 US-09-854-133-145	Sequence 145, App
27	74	3.7	114793	4 US-10-148-806-3	Sequence 3, Appl1

28	70.2	3.5	215	4 US-09-513-999C-31510	Sequence 31510, A
29	67	3.3	152132	4 US-09-949-016-13885	Sequence 13885, A
30	67	3.3	152145	4 US-09-949-016-12371	Sequence 12371, A
31	66.4	3.3	311	4 US-09-313-294A-5928	Sequence 5928, Ap
32	65.6	3.2	7218	1 US-08-232-463-14	Sequence 14, Appl
33	64.6	3.2	601	4 US-09-949-016-176640	Sequence 176640, A
34	64.6	3.2	601	4 US-09-949-016-176663	Sequence 176663, A
35	60	3.0	268	1 US-08-437-027-19	Sequence 19, Appl
36	59.8	3.0	1505	1 US-07-915-246-1	Sequence 1, Appl1
37	59	2.9	1225	4 US-09-976-594-416	Sequence 416, A
38	58.8	2.9	34230	4 US-09-949-016-12052	Sequence 12052, A
39	58.8	2.9	128470	4 US-09-949-016-13765	Sequence 13765, A
40	58.4	2.9	425	2 US-08-343-443B-45	Sequence 45, Appl
41	58	2.9	328	2 US-08-343-443B-5	Sequence 5, Appl1
42	57.8	2.9	12695	4 US-09-949-016-16775	Sequence 16775, A
43	57.2	2.8	1682	4 US-09-220-132-82	Sequence 82, Appl
44	57.2	2.8	1684	4 US-09-919-039-323	Sequence 33, App
45	57	2.8	1926	3 US-09-249-585A-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/0834343B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:







Db	3681	GGCTTCTCCCGCACCTTCCACCCCGGGGTGTATCATGGCAGAGGTGGCCCTGGT	3740
QY	1744	GGCATCGGGGAGGAGNAGGTGGCTCATGGAATGCTGTGCTCCCGGTGAAATTGAGA	1803
Db	3741	GGCATGTGGGAGGAAAGGTGGCTCATGATCATGTGTGCTCCCGGTGAATGTTCAGA	3800
QY	1804	GGTGGCGGTGTGGAGACAGAGGTGGCTCCGTGTGTGCGGGGGCATGTGAACGAGGTGGC	1863
Db	3801	GGTGCCTGTGTAGAGACAGAAAGTGGCTCCGTGTGTG-CTGGGCAATGACAGAGGTGGC	3859
QY	1864	TTTGTGTGAGAGACGAGGTGGCTTCCTGTGGGGGGCCCCCTGTGACCTTTGTATGAAACAGATG	1923
Db	3860	TTTGTGTGAGAAAGCAAGGTGGCTTCCTGTGGGGGGCCCC-GGACCTTTGATTAACAATG	3918
QY	1924	GGAGGAAAGAGAGACGTGTGAGAACCTTGGAAAAATGATTTAAAGGTGACACCTGTACG	1983
Db	3919	GGAGGAAAGAGAGACGTGTGAGAACCTTGGAAAAAGGATTAAGGCGAGACCTGTTCAG	3978
QY	1984	GAGGCGAGAGATCGGCTCTACTAAGTGCAGAGACCCCGCAGA	2025
Db	3979	GAGGCGAGAGATCAGCTCTACTAAGTGCAGAGAACCCCGCAGA	4020

### RESULT 3

```

US-09-949-016-5043
; Sequence 5043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5043
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5043

```

Query Match	76.8%	Score 1556;	DB 4;	Length 1785;
Best Local Similarity	89.1%	Pred. No. 0;		
Matches 1785; Conservative	0;	Mismatches 0;	Indels 219;	Gaps 1;

QY	4	GGACGTTAGAGAAAGAAAGAGAGAAATGGGGGTCCAGGATTAAGTACCTAT	63
Db	1	GGACCTTAGAGAAACGAGAGGAAAGAGAAATGGGGTCCAGGATTAAGTACCTAT	60
QY	64	AGCCAAAGTTCAGCGCAGACGGGCTTACAGTCTTACACCGCCAGCCCACTTCAGAGATAT	123
Db	61	AGCCAAAGTTCAGCGCCAGACGGGCTTACAGTCTTACACCGCCCAAGCTTCAGAGATAT	120
QY	124	GCACGACCAACCCAGGCAATATGGGGCAACAAAGCTATGAGAACTTATGACAGGCCATAT	183
Db	121	GCACGACCAACCCAGGCAATATGGGGCAACAAAGCTATGAGAACTTATGACAGGCCATAT	180
QY	184	GTCAAGCTATACCCAGGCTCAGACCACTGACCACTTATGGGAGAACCGGCTATGCACTTCT	243
Db	181	GTCAAGCTATACCCAGGCTCAGACCACTGACCACTTATGGGAGAACCGGCTATGCACTTCT	240
QY	244	TATGAGACAGCTCCCACTGTTATATCTCAACTGGCCCCCAGGCAATACAGCCAGCT	303
Db	241	TATGAGACAGCTCCCACTGTTATATCTCAACTGGCCCCCAGGCAATACAGCCAGCT	300

QY	304	GTCAAGGGGATATGGCACTGGGGCTTATGATACCAACCACTGCTCAGTCAACCAACCCAG	363
Db	301	GTCCAGGGGATATGGCACTGGGGCTTATGATACCAACCACTGCTCAGTCAACCAACCCAG	360
QY	364	GCTCTCTATGACAGCTCAGTCTGCATATATGGACTCAGCGTCTTATCCAGGCTTATGGCAG	423
Db	361	GCTCTCTATGAGTCAAGTCTGCATATATGGACTCAGCGTCTTATCCAGGCTTATGGCAG	420
QY	424	CAGCCAGACGCACTGCACCTTACAAAGCCGAGATGGAAACAAGCCCACTGAGACTATAGT	483
Db	421	CAGCCAGACGCACTGCACCTTACAAAGCCGAGATGGAAACAAGCCCACTGAGACTATAGT	480
QY	484	CAACCTCAATATGACAGAGGGGGTTAAACAGGCCAGCTAGGATATGACAGAGTAAAC	543
Db	481	CAACCTCAATATGACAGAGGGGGTTAAACAGGCCAGCTAGGATATGACAGAGTAAAC	540
QY	544	TACAGTTATCCCGAGTATCTGGAGAGCTACCCCATGAGCCAGTCACTGCACCTCCATCC	603
Db	541	TACAGTTATCCCGAGTATCTGGAGAGCTACCCCATGAGCCAGTCACTGCACCTCCATCC	600
QY	604	TACCTCTCTACAGCTATTCCTCTTACACAGCCGACTATGATATGATCAGACAGTTACTCT	663
Db	601	TACCTCTCTACAGCTATTCCTCTTACACAGCCGACTATGATATGATCAGACAGTTACTCT	660
QY	664	CAGAGAAACATATATGGGCAACCGAGAGACTATGGAACAGACAGTACTATGGTCACAA	723
Db	661	CAGAGAAACATATATGGGCAACCGAGAGACTATGGAACAGACAGTACTATGGTCACAA	720
QY	724	AGCAGCTATATGGGCGACAGCTCCCACTAGTTATCCCACTGGATCTTACAGCCAA	783
Db	721	AGCAGCTATATGGGCGACAGCTCCCACTAGTTATCCCACTGGATCTTACAGCCAA	780
QY	784	GCTCCAAATCAATATAGCCAAACAGACAGCACTACGGGCGACAGAGTTCAATCCGACAG	843
Db	781	GCTCCAAATCAATATAGCCAAACAGACAGCACTACGGGCGACAGAGTTCAATCCGACAG	826
QY	844	GACCAACCCAGTACATATGGGTGTTATATGGGAGAGTCTGGAGGATTTTCCGACACAGA	903
Db	827	-----	826
QY	904	GAGAACCGGAGCATGATGAGGCTCTGATTAACCGGGGAGAGGGAAGGGGAGTTGATCGT	963
Db	827	-----	826
QY	964	GGAAGCATGACAGAGGTGGCGGGGAGAGAGACGCGGTGGAATGGGCAGCGCTGAGAG	1023
Db	827	-----	826
QY	1024	GGAAGTGGCTTCAATTAAGCTGTGGACCCATGGATTAAGAACAGATCTTGAATCTTAGGC	1083
Db	827	-----GACCCATGGATTAAGAACAGATCTTGAATCTTAGGC	861
QY	1084	CCACCTGATATCCAGATGAGAGCTCTGCACAAAGTGCATTTTATGTAACAAGATTAAAT	1143
Db	862	CCACCTGATATCCAGATGAGAGCTCTGCACAAAGTGCATTTTATGTAACAAGATTAAAT	921
QY	1144	GACAGTGTGACTTATGATGATCTTGGCAGACTTCTTTAAGCAGTGTGGGGTGTTAAGATG	1203
Db	922	GACAGTGTGACTTATGATGATCTTGGCAGACTTCTTTAAGCAGTGTGGGGTGTTAAGATG	981
QY	1204	AACAAGAGAACTTGGGCAACCCATGATCCAACTTACTCTGACACAGAAACAAGAAACCC	1263
Db	982	AACAAGAGAACTTGGGCAACCCATGATCCAACTTACTCTGACACAGAAACAAGAAACCC	1041
QY	1264	AAAGGCGATGCAAGATGCTCTATGAAGACCACCACTGCAAGGCTGCGGTGGAATGG	1323
Db	1042	AAAGGCGATGCAAGATGCTCTATGAAGACCACCACTGCAAGGCTGCGGTGGAATGG	1101
QY	1324	TTTGTATGGAAAGATTTTCAAGGAGCAACTTAAAGTCTTCCCTTGTGGAAGAGCT	1383
Db	1102	TTTGTATGGAAAGATTTTCAAGGAGCAACTTAAAGTCTTCCCTTGTGGAAGAGCT	1161
QY	1384	CCAAATGAACAGTATGCGGGGTGTCTGCAACCCCGTATGGGCAAGAGCATGCAACACCA	1443

Db 1162 CCAATGAACATATGCGGGGTGTCTGCGAACCCCTGAGGGCAGAGGCTATCCACCA 1221  
Qy 1444 CTCCTGAGAGGTCCAGAGGCCAGAGGCTCTGAGGAGCCCATGAGTCCGATGAGAGGC 1503  
Db 1222 CTCCTGAGAGGTCCAGAGGCCAGAGGCTCTGAGGAGCCCATGAGTCCGATGAGAGGC 1281  
Qy 1504 CTTGAGAGAGATTAAGAGGCTTCTCTCAAGAGAGCCCGGGGTTCCTCGAGGAGAACCC 1563  
Db 1282 CTTGAGAGAGATTAAGAGGCTTCTCTCAAGAGAGCCCGGGGTTCCTCGAGGAGAACCC 1341  
Qy 1564 TCTGAGAGAGAGATTCAGACCCAGAGCTGAGAGCTGAGAGTCCCAATCCGGATTGT 1623  
Db 1342 TCTGAGAGAGAGATTCAGACCCAGAGCTGAGAGCTGAGAGTCCCAATCCGGATTGT 1401  
Qy 1624 GGAACCAAGAACTTCCCTGAGAAACAGAGTGAACCACTGTAAGGCCCCAAAGCTGAA 1683  
Db 1402 GGAACCAAGAACTTCCCTGAGAAACAGAGTGAACCACTGTAAGGCCCCAAAGCTGAA 1461  
Qy 1684 GCTTCTCCCGCAACCTTCCCGCCCGGGTGTATCTGGCAGAGGTGGCCCTGTGT 1743  
Db 1462 GCTTCTCCCGCAACCTTCCCGCCCGGGTGTATCTGGCAGAGGTGGCCCTGTGT 1521  
Qy 1744 GGCATGCGGGAGAGAGAGGTGGCTCATGATCTGGTGGTCCCGTGAATGTTGAA 1803  
Db 1522 GGCATGCGGGAGAGAGAGGTGGCTCATGATCTGGTGGTCCCGTGAATGTTGAA 1581  
Qy 1804 GGTGCGCTGTGTGAGACAGAGTGTCTCCGTGTGTGGCCGGGCGATGACCGAGTGGC 1863  
Db 1582 GGTGCGCTGTGTGAGACAGAGTGTCTCCGTGTGTGGCCGGGCGATGACCGAGTGGC 1641  
Qy 1864 TTTGTGTGAGAGAGACAGAGTGTGGCTTGGGGGGCCCTTGGACCTTTGATGAAACAGATG 1923  
Db 1642 TTTGTGTGAGAGAGACAGAGTGTGGCTTGGGGGGCCCTTGGACCTTTGATGAAACAGATG 1701  
Qy 1924 GAG 1983  
Db 1702 GAG 1761  
Qy 1984 GAGCGCAGAGATCGGCTCTACTAG 2007  
Db 1762 GAGCGCAGAGATCGGCTCTACTAG 1785

RESULT 4  
US-09-949-016-5044  
Sequence 5044, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949, 016  
PRIORITY FILING DATE: 2000-04-14  
PRIORITY APPLICATION NUMBER: 60/241,755  
PRIORITY FILING DATE: 2000-10-20  
PRIORITY APPLICATION NUMBER: 60/237,768  
PRIORITY FILING DATE: 2000-10-03  
PRIORITY APPLICATION NUMBER: 60/231,498  
PRIORITY FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5044  
LENGTH: 1785  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-5044

Query Match 76.8%; Score 1556; DB 4; Length 1785;  
Best Local Similarity 89.1%; Pred. No. 0;  
Matches 1785; Conservative 0; Mismatches 0; Indels 219; Gaps 1;

Qy 4 GAGCTTGAAGAAACGAG 63  
Db 1 GAGCTTGAAGAAACGAG 60  
Qy 64 AGCCAGCTCAGGCGAT 123  
Db 61 AGCCAGCTCAGGCGAT 120  
Qy 124 GCACAGACCAACCGAGATATGAGCAACAAAGCTATGAAACCTATGAGACAGCCACTGAT 183  
Db 121 GCACAGACCAACCGAGATATGAGCAACAAAGCTATGAAACCTATGAGACAGCCACTGAT 180  
Qy 184 GTGAGCTATACCAAGGCTCAGACCACTGCAACCTATGAGAGAGAGAGAGAGAGAGAG 243  
Db 181 GTGAGCTATACCAAGGCTCAGACCACTGCAACCTATGAGAGAGAGAGAGAGAGAGAG 240  
Qy 244 TATGAGAGAGCTCCCACTGATTAATCACTCACTCCCGGAGAGATGAGAGAGAGAGAG 303  
Db 241 TATGAGAGAGCTCCCACTGATTAATCACTCACTCCCGGAGAGATGAGAGAGAGAGAG 300  
Qy 304 GTCCAGGGGATATGAGCACTGATGCTTATGATACCAACCTGCTACAGTCAACCAACCCAG 363  
Db 301 GTCCAGGGGATATGAGCACTGATGCTTATGATACCAACCTGCTACAGTCAACCAACCCAG 360  
Qy 364 GCTTCTTATGAGCTGATGCTGATATGAGCACTGAGCTGATGCTTATGAGAGAGAGAG 423  
Db 361 GCTTCTTATGAGCTGATGCTGATATGAGCACTGAGCTGATGCTTATGAGAGAGAGAG 420  
Qy 424 CAGCAG 483  
Db 421 CAGCAG 480  
Qy 484 CAACCTCAATCTAGCAACAGGGGTTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 543  
Db 481 CAACCTCAATCTAGCAACAGGGGTTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Qy 544 TACAGTTATCCCAAGTATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603  
Db 541 TACAGTTATCCCAAGTATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Qy 604 TACCTCTCAAGAGCTATCTCTCTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663  
Db 601 TACCTCTCAAGAGCTATCTCTCTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Qy 664 CAGCAGAGACCTATGAGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 723  
Db 661 CAGCAGAGACCTATGAGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Qy 724 AGCAGATATGAG 783  
Db 721 AGCAGATATGAG 780  
Qy 784 GCTCAAGTATATGAG 843  
Db 781 GCTCAAGTATATGAG 840  
Qy 844 GACCAACCCAGTACATGAGTGTGTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 903  
Db 842 GACCAACCCAGTACATGAGTGTGTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Qy 904 GAGAACCGAGAGATGAGTGGCTCTGATACCGGGGAGAGAGAGAGAGAGAGAGAGAT 963  
Db 827 GAGAACCGAGAGATGAGTGGCTCTGATACCGGGGAGAGAGAGAGAGAGAGAGAGAT 826  
Qy 964 GAGAGATGAG 1023  
Db 827 GAGAGATGAG 826  
Qy 1024 CAGAGGTGCTTCAATAGAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1083  
Db 827 CAGAGGTGCTTCAATAGAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 861  
Qy 1084 CAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143

Db	862	CCACGTAGATCCAGATGAGACCTCTGCACAACAGTCAATTTATGTACAGGAATTAAAT	921
Oy	1144	GACAGTGTGACTCTAGATGATCTGGCAGACTTTTAAACAGTGTGGGTTGTTAAGATG	1203
Db	922	GACAGTGTGACTCTAGATGATCTGGCAGACTTTTAAACAGTGTGGGTTGTTAAGATG	981
Oy	1204	AACAAGAAACTGGGCAACCCATGATCCATCTACCTGACACAGAAACAGAAAGCCC	1263
Db	982	AACAAGAAACTGGGCAACCCATGATCCATCTACCTGACACAGAAACAGAAAGCCC	1041
Oy	1284	AAAGCGATGCCACAGTGTCTTATGAAAGCCACCACTGCCAAGGCTGCCTGAAATGG	1323
Db	1042	AAAGCGATGCCACAGTGTCTTATGAAAGCCACCACTGCCAAGGCTGCCTGAAATGG	1101
Oy	1324	TTGATGAGGAAGATTTTCAAGGGAGCAAACTTAAAGTCCCTCTGCGAAAGAACCT	1383
Db	1102	TTGATGAGGAAGATTTTCAAGGGAGCAAACTTAAAGTCCCTCTGCGAAAGAACCT	1161
Oy	1384	CCAAATGAACATGATGCGGGTGTCTGCCACCCCGTGAAGGACAGAGCATGCCACCA	1443
Db	1182	CCAAATGAACATGATGCGGGTGTCTGCCACCCCGTGAAGGACAGAGCATGCCACCA	1221
Oy	1444	CTCCGTGAGGTCCAGAGAGCCCAAGAGGTCTTGGGGAGACCATGGGTGCATGGAGGC	1503
Db	1222	CTCCGTGAGGTCCAGAGAGCCCAAGAGGTCTTGGGGAGACCATGGGTGCATGGAGGC	1281
Oy	1504	CGTGAAGAGATAGAGGAAGCTTCCCTCCAAAGAGACCCCGGGGTCCCGAGGAAACCC	1563
Db	1282	CGTGAAGAGATAGAGGAAGCTTCCCTCCAAAGAGACCCCGGGGTCCCGAGGAAACCC	1341
Oy	1564	TCTGAGAGAGAAACGTCCAGACCCGAGCTGAGACATGACAGTCCCAATCCGGATTGT	1623
Db	1342	TCTGAGAGAGAAACGTCCAGACCCGAGCTGAGACATGACAGTCCCAATCCGGATTGT	1401
Oy	1624	GGAACACCAACTCTTGCTCTGAGAAACAGAGTGCACCACTGTAAAGCCCCAAAGCTTGA	1683
Db	1402	GGAACACCAACTCTTGCTCTGAGAAACAGAGTGCACCACTGTAAAGCCCCAAAGCTTGA	1461
Oy	1684	GGCCTTCCTCCGCGCACCTTTCCGAGCCCGGGGTGTGATCGTGGCAGAGGTGACCCTGGT	1743
Db	1462	GGCCTTCCTCCGCGCACCTTTCCGAGCCCGGGGTGTGATCGTGGCAGAGGTGACCCTGGT	1521
Oy	1744	GGCATGCGGAGAGAAAGAGTGGCTCATGATCGTGTGTCTCCGCTGGAAATGTTCAAG	1803
Db	1522	GGCATGCGGAGAGAAAGAGTGGCTCATGATCGTGTGTCTCCGCTGGAAATGTTCAAG	1581
Oy	1804	GGTGGCCCTGTGTGAGACACAGGTGGCTTCCGTGTGTGCGCGGGCAATGACCGAGGTGGC	1863
Db	1582	GGTGGCCCTGTGTGAGACACAGGTGGCTTCCGTGTGTGCGCGGGCAATGACCGAGGTGGC	1641
Oy	1864	TTTGTGTGAGAAAGACAGAGTGGCTCTGGGGGGCCCCCTTGAGACTTTTGAATGAAACAGATG	1923
Db	1642	TTTGTGTGAGAAAGACAGAGTGGCTCTGGGGGGCCCCCTTGAGACTTTTGAATGAAACAGATG	1701
Oy	1924	GGAGGAAGAAAGAGAGACGTGAGAGACCTGGAAGAAATGATTAAGCGAGCACCTGTAG	1983
Db	1702	GGAGGAAGAAAGAGAGACGTGAGAGACCTGGAAGAAATGATTAAGCGAGCACCTGTAG	1761
Oy	1984	GAGCCGAGATCGGCGCTTAATG 2007	
Db	1762	GAGCCGAGATCGGCGCTTAATG 1785	

RESULT 5  
 US-09-949-016-1954  
 ; Sequence 1954, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J Craig et al  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307

	CURRENT APPLICATION NUMBER:	US-09/949, 016	
	PRIOR FILING DATE:	2000-04-14	
	PRIOR APPLICATION NUMBER:	60/241, 755	
	PRIOR FILING DATE:	2000-10-20	
	PRIOR APPLICATION NUMBER:	60/237, 768	
	PRIOR FILING DATE:	2000-10-03	
	PRIOR APPLICATION NUMBER:	60/231, 498	
	PRIOR FILING DATE:	2000-09-08	
	NUMBER OF SEQ ID NOS:	207012	
	SOFTWARE:	FastSeq for Windows Version 4.0	
	SEQ ID NO:	1954	
	LENGTH:	1783	
	TYPE:	DNA	
	ORGANISM:	Human	
	US-09-949-016--1954		
	Query Match	70.6%; Score 1429.6; DB 4; Length 1783;	
	Best Local Similarity	85.8%; Pred. No. 0;	
	Matches 1719; Conservative	0; Mismatches 64; Indels 221; Gaps 3	
QY	4	GGAGCTTGAGAGAAACGAGAGAGAAAGAGAGAAAATGGGTGCCAGGATTACAGTACTCAT	63
DB	1	GGAAGTTGAGAGAAACGAGAGAGAGAGAAAAATGGATCCTAATGATTACAGTACTCAT	60
QY	64	AGCCAAAGTCGAGGCGACGAGGCGCTCAAGTCTTAACAACGCGCCAGCCCAAGATAT	123
DB	61	AGCCAAAGTCGAGGCGACGAGGCGCTCAAGTCTTAACAACCAGCCCACTCAAGATAT	120
QY	124	GCAAGAACCAACCCAGGCATATGAGCAACAAAGCTATGGAACCTATGACAGGCCATGAT	183
DB	121	GCAAGAACCAACCCAGGCATATGAGCAACAAAGCTATGGAACCTATGACAGGCCATGAT	180
QY	184	GTCAGCTATACCCAGGCTCAGAACAATGCAACTATATGAGGACAGCCGCTTATGCAACTTCT	243
DB	181	GTCAGCTATACCCAGGCTCAGAACCACTGCAATCTATGAGGACAGCCGCTTATGCAACTTCT	240
QY	244	TATGAGACAGCCTCCCACTGGTTATTAATACTCAACTGGCCCCGAGGATACAGCAGCCT	303
DB	241	TATGAGACAGCCTCCCACTGGTTATTAATACTCAACTGGCCCCGAGGATACAGCAGCCT	300
QY	304	GTCCAAGGGGTATGSCACTGTGTCTTATGATACCAACAATGCTATCAGTACCAACCAACCGAG	363
DB	301	GTCCAAGGGGTATGSCACTGTGTCTTATGATACCAACAATGCTATCAGTACCAACCAACCGAG	360
QY	364	GCTTCTTATGCAAGCTCAAGTCTGCAATATGGAATGCACTGAGCCTGCTTATTCAGCCTTATGGGCAAG	423
DB	361	GCTTCTTATGCAAGCTCAATCTGCAATATGGAATGCACTGAGCCTGCTTATTCAGCCTTATGGGCAAG	420
QY	424	CAGCAGAGAGCACTGCACTTAACAAGACCGAGAGATGGAACCAAGCCCACTGAGACTAGT	483
DB	421	CAGCAGAGAGCACTGCACTTAACAAGACCGAGAGATGGAACCAAGCCCACTGAGACTAGT	480
QY	484	CAACCTCAATCTAGACAAGGGGGTTTACAAACCAAGCCCACTAGATATGACAGAGTAAC	543
DB	481	CAACCTCAATCTAGACAAGGGGGTTTACAAACCAAGCCCACTAGATATGACAGAGTAAC	540
QY	544	TACAGTTATCCCAGGTACTTGGGAGCTAACCCCATGACGCCAGTCACTGCACTCTCATCC	603
DB	541	TGCAAGTTATCCCAGGTACTTGGGAGCTAACCTCATGACGCCAGTCACTGCACTCTCATCC	600
QY	604	TACCTCTTACAGCTATTTCTCTTACAAGCGCACTAATTATGATCAGAGCAAGTTACTCT	663
DB	601	TACCTCTTACAGCTATTTCTCTTACAAGCGCACTAATTATGATCAGAGCAAGTTACTCT	660
QY	664	CAGCAGAACACTATATGAGCAACCGAGAGCTATGGAACAGAGATGATGTTCAACA	723
DB	661	CAGCAGAACACTATATGAGCAACCGAGAGCTATGGAACAGAGATGATGTTCAACA	720
QY	724	AGCAGCTATGAGGACAGCAGCTTCCAATTATGTTACCAACCCCAAATGATCTTATCAGCCAA	783
DB	721	AGCAGCTATGAGGACAGCAGCTTCCAATTATGTTACCAACCCCAAATGATCTTATCAGCCAA	780
QY	784	GCTCCAAAGTCAATATAGCCAAACAGAGCACAGCTACGGGAGCAGACAGTTCAATTCGACAG	843



QY	253	CGTCCACGCTGGTAACTACTCTCAACTGCGCCCCCAGGAGATACAGGACGCTGTGCAGGGG	3112
Db	241	CTCTCCACTGGTTATATCTACTCCACTGCGCCCCCAGGAGATACAGGACGCTGTGCAGGGG	3000
QY	313	TATGGCACTGGTGCTTATATGATACCACTGTGTACAGTACCAACCACCGAGCCTCTTAT	3722
Db	301	TATGGCACTGGTGCTTATATGATACCACTGTGTACAGTACCAACCACCGAGCCTCTTAT	3600
QY	373	GCAGCTCAGCTGCGAATATGGCACTGACGCTGTGTATCCAGGCTATATGGGACAGCCAGCA	4322
Db	361	GCAGCTCAGCTGCGAATATGGCACTGACGCTGTGTATCCAGGCTATATGGGACAGCCAGCA	4200
QY	433	GCGACTGACCTTACAAAGACCGAGATGGAAACAAGCCCATGAGACTAGTCAACTCAA	4922
Db	421	GCGACTGACCTTACAAAGACCGAGATGGAAACAAGCCCATGAGACTAGTCAACTCAA	4800
QY	493	TCTAGCAGAGGGGGTTACAAACAGCCGACCTAGGATATGACAGAGTAACTACAGTTAT	5522
Db	481	TCTAGCAGAGGGGGTTACAAACAGCCGACCTAGGATATGACAGAGTAACTACAGTTAT	5400
QY	553	CCCCAGGTACTCTGGAGGCTTACCCCATGAGCCAGTCACTGCACTCTCCATCCTAACCTTCCT	6122
Db	541	CCCCAGGTACTCTGGAGGCTTACCCCATGAGCCAGTCACTGCACTCTCCATCCTAACCTTCCT	6000
QY	613	ACGACGATATTCCTCTACACAGCCGACTATATGATCAGAGCAGTTACTCTCAGACAAAC	6722
Db	601	ACGACGATATTCCTCTACACAGCCGACTATATGATCAGAGCAGTTACTCTCAGACAAAC	6600
QY	673	ACCTATGGGGACCCGAGCAGCTATGAGACAGCAGAGTACGTATGTCAACAAAGCAGCTAT	7322
Db	661	ACCTATGGGGACCCGAGCAGCTATGAGACAGCAGAGTATGTGTCAACAAAGCAGCTAT	7200
QY	733	GGGAGAGAGGCTCCCACTAGTTATCCCAACCCCAATCGATTCCTAAGCCAAAGCTCCAGT	7922
Db	721	GGGAGAGAGGCTCCCACTAGTTATCCCAACCCCAATCGATTCCTAAGCCAAAGCTCCAGT	7800
QY	793	CAATATAGCCAAAGAGCAGCAGCTATCGGGCAGCAGAGT 831	
Db	781	CAATATAGCCAAAGAGCAGCAGCTATCGGGCAGCAGAGT 819	

```

      QY          1 GACGAGCTTGAGAGAACGAGGAGGAGGAAATGGCGTCCAGCATTCAGTACC 60
|||||
US-09-513-999C-1657
RESULT 7
Sequence 1657; Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumae Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59. US2.REG
CURRENT APPLICATION NUMBER: US/09/513, 999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 1657
LENGTH: 454
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 56..454
US-09-513-999C-1657
Query Match
Best Local Similarity 99.8%; Score 424; DB 4; Length 454;
Matches 435; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Db	20	GACGGACGTTAGAGAAACAGAGAGGAAAGAGAAATGGCGTCCACGGATTACAGTACC	79
Qy	61	TATAGCCAGGCTGCAGCGCAGAGGSGCTTACAGTGGTTACACCGGCCAGCCCTCACAAGA	120
Db	80	TATAGCCAGGCTGCAGCGCA - CAGGSGCTACAGTGGCTTACCGGCCAGCCCACTCACAAGA	138
Qy	121	TATGCACAGACCAACCAGGCGATATGGGGCAACAAAGCTATGAAACCTATGTGAACAGCCCACT	180
Db	139	TATGCACAGACCAACCAGGCGATATGGGGCAACAAAGCTATGAAACCTATGGAAGCCCACT	198
Qy	181	GATGTCAGCTATACCCAGGCTCAGACCACTGAAACTGTAATGGGCAAGCGGCTTATGCAACT	240
Db	199	GATGTCAGCTATACCCAGGCTCAGACCACTGCAACTGTAATGGGCAAGCGGCTTATGCAACT	258
Qy	241	TCTTATGACAGCCCTCCACTGTTATCTACTCCAACTGGCCCCCAGGCAATACAGCCAG	300
Db	259	TCTTATGACAGCCCTCCACTGTTATCTACTCCAACTGGCCCCCAGGCAATACAGCCAG	318
Qy	301	CCTGTCAGGGGATATGGCACTGATGCTTATGATATACCAACACTGCTACAGTACCAACACC	360
Db	319	CCTGTCAGGGGATATGGCACTGATGCTTATGATATACCAACACTGCTACAGTACCAACACC	378
Qy	361	CAGGCGCTCTATGACAGCTCAGTCTGCATATGGCACTCAGCGCTGCTTATCCAGGCTTATGGG	420
Db	379	CAGGCGCTCTATGACAGCTCAGTCTGCATATGGCACTCAGCGCTGCTTATCCAGGCTTATGGG	438
Qy	421	CAGCAGCCAGCAGCCA	436
Db	439	CAGCAGCCAGCAGCCA	454

```

RESULT 8
US-09-949-016-66382/c
; Sequence 66382, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 66382
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-66382

Query Match      18.0%; Score 364; DB 4; Length 601;
Best Local Similarity 98.1%; Pred. No. 4.3e-88;
Matches 367; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

4 GCACGTTGAGAGACGAGGAGAGAGAGAAATAGCGCTCCACGGATTACAGTACTTAT 63
Db 374 GGACCGTTGAGAGACGAGGAGAGAGAGAAATAGCGCTCCATGATTACAGTACTTAT 315

64 AGCCAGCTGCGAGCCAGAGAGGGCTTACAGTGGCTTACACGCCCGCCAGCCACTCAAGATAT 123
Db 314 AGCCAGCTGCGAGGAGAGAGGGCTTACAGTGGCTTACACGCCCGCCACTCAAGATAT 255

124 GCACAGACCAACCAGGCAATATGGGCAACAAGCTATGAACTTATGAGACGCCCACTGAT 183
Db 254 GCACAGACCAACCAGGCAATATGGGCAACAAGCTATGAACTTATGAGACGCCCACTGAT 195

184 GTACAGTATCCACAGGCTCAGACCACTGCACACTGACACTTATGGGCAAGCCGCTTATGCAACTTCT 243

```





QY 121 TATGACAGACCAACCCAGGATATGGGCAACAAAGCTATGGAACCTATGACAGCCACT 180  
DB 128 TATGACAGACCAACCCAGGATATGGGCAACAAAGCTATGGAACCTATGACAGCCACT 187  
QY 181 GATGTGAGTATACCCAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGCAACT 240  
DB 188 GATGTGAGTATACCCAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGCAACT 247  
QY 241 TCTTATGAGACGCTCCCACTGTTTATATCTCAACCTGCCCCCGGATACAGCTAG 300  
DB 248 TCTTATGAGACGCTCCCACTGTTTATATCTCAACCTGCCCCCGGATACAGCTAG 307  
QY 301 CCGTCCAGGGGATATGGGCACTGGGCTTATGATACCACTGCTTACAGTACCAACACC 360  
DB 308 CCGTCCAGGGGATATGGGCACTGGGCTTATGATACCACTGCTTACAGTACCAACACC 367  
QY 361 CAGGCTCTTATGAGCTCAGTCTGCAATATGGCACTGAGCTGTTATCCAGCTATGGG 420  
DB 368 CAGGCTCTTATGAGCTCAGTCTGCAATATGGCACTGAGCTGTTATCCAGCTATGGG 427  
QY 421 CAGGACGACAGCCACTGCACTCAACAGACCCGAGATGAAACAAAGCCCACTGAGACT 480  
DB 428 CAGGACGACAGCCACTGCACTCAACAGACCCGAGATGAAACAAAGCCCACTGAGACT 487  
QY 481 AGTCAACCTCAATCTAGACAGAGGGGTTTACAAACAGCCAGCCCTAGGATATGACAG 540  
DB 488 AGTCAACCTCAATCTAGACAGAGGGGTTTACAAACAGCCAGCCCTAGGATATGACAG 547  
QY 541 AACTACAGTTATCCCAAGTATCCTGAGAGCTACCCCACTGCAAGCCAGTCACTGCACTCA 600  
DB 548 AACTACAGTTATCCCAAGTATCCTGAGAGCTACCCCACTGCAAGCCAGTCACTGCACTCA 607  
QY 601 TCTCACTCTCTACAGCTATCTCTCTACACAGCCGCTAGTTATGATCAGAGAGTTAC 660  
DB 608 TCTCACTCTCTACAGCTATCTCTCTACACAGCCGCTAGTTATGATCAGAGAGTTAC 667  
QY 661 TCTCAGAGAAACCTATGGGCAACCGAGAGCTATGAGAGAGTATGAGTATGCTCA 720  
DB 668 TCTCAGAGAAACCTATGGGCAACCGAGAGCTATGAGAGAGTATGAGTATGCTCA 727  
QY 721 CAAAGCAGTATGGGCAAGCCCTCCCACTAGTTACCAACCCCACTGATCTCAAGC 780  
DB 728 CAAAGCAGTATGGGCAAGCCCTCCCACTAGTTACCAACCCCACTGATCTCAAGC 787  
QY 781 CAAGCTCAAGTCAATATGACCAACAGAGCAGCTACCGGAGCAGAGTTCATTCGA 840  
DB 788 CAAGCTCAAGTCAATATGACCAACAGAGCAGCTACCGGAGCAGAGTTCATTCGA 847  
QY 841 CAGGACCAACCCAGTACATGGGCTGTTTATGGGAGAGTCTGAGAGATTTTCCGAGCA 900  
DB 848 CAGGACCAACCCAGTACATGGGCTGTTTATGGGAGAGTCTGAGAGATTTTCCGAGCA 907  
QY 901 GAGAGAAACCGGAGCATGAGTGGCCCTGATTAACCGGGGCAAGGGAGATTTTAT 960  
DB 908 GAGAGAAACCGGAGCATGAGTGGCCCTGATTAACCGGGGCAAGGGAGATTTTAT 967  
QY 961 CGTGAGAGCATGAGAGAGTGGGCGGGAGAGAGACCGGCTGAAATGGGAGCGCTGGA 1020  
DB 968 CGTGAGAGCATGAGAGAGTGGGCGGGAGAGAGACCGGCTGAAATGGGAGCGCTGGA 1027  
QY 1021 GAGGAGAGTGGCTCAATTAAGCTGATGAGACCCATGATGAGAGACCATCTGATCTA 1080  
DB 1028 GAGGAGAGTGGCTCAATTAAGCTGATGAGACCCATGATGAGAGACCATCTGATCTA 1087  
QY 1081 GGGCCACCTGATGATGATGAGAGACTCTGACAAACAGTGAATTTATGATCAAGATTA 1140  
DB 1088 GGGCCACCTGATGATGATGAGAGACTCTGACAAACAGTGAATTTATGATCAAGATTA 1147  
QY 1141 AATGACAGTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
DB 1148 AATGACAGTGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1207  
QY 1201 ATGAACAAGAGAACTGGGCAACCATGATGATGATGATGATGATGATGATGATGATGAT 1260

DB 1208 ATGAACAAGAGAACTGGGCAACCATGATGATGATGATGATGATGATGATGATGATGAT 1267  
QY 1261 CCCAAAGGAGATGCCAGTGTCTTATGAAAGCCCACTGCAAGAGTGTGCGTGGAA 1320  
DB 1268 CCCAAAGGAGATGCCAGTGTCTTATGAAAGCCCACTGCAAGAGTGTGCGTGGAA 1327  
QY 1321 TGGTTGATGGGAAATTTTCAAGGAGCAACTTAAAGTCTCCCTGTGCTGGAGAAAG 1380  
DB 1328 TGGTTGATGGGAAATTTTCAAGGAGCAACTTAAAGTCTCCCTGTGCTGGAGAAAG 1387  
QY 1381 CCTCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGCATGCCCA 1440  
DB 1388 CCTCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGAGCATGCCCA 1447  
QY 1441 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCTGCGGGAGACCATGGTCCGATGGGA 1500  
DB 1448 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCTGCGGGAGACCATGGTCCGATGGGA 1507  
QY 1501 GGCCTGAGAGAGATGAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCGAGGGAA 1560  
DB 1508 GGCCTGAGAGAGATGAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCGAGGGAA 1567  
QY 1561 CCCTTGAAGAGAGAAACGTCCAGACCGAGCTGGAAGCTGAGAGTGTCCCAATCCGGGT 1620  
DB 1568 CCCTTGAAGAGAGAAACGTCCAGACCGAGCTGGAAGCTGAGAGTGTCCCAATCCGGGT 1627  
QY 1621 TGTGAAACCAAGAACTTCCCTGAGAAACAGAGTGTCAACAGTGTAAAGGCCCAAGGCT 1680  
DB 1628 TGTGAAACCAAGAACTTCCCTGAGAAACAGAGTGTCAACAGTGTAAAGGCCCAAGGCT 1687  
QY 1681 GAAAGCTTCTCCCGGCAACCTTCCGCGCCCGGGGTGTGATGATGATGATGATGATGAT 1740  
DB 1688 GAAAGCTTCTCCCGGCAACCTTCCGCGCCCGGGGTGTGATGATGATGATGATGATGAT 1747  
QY 1741 GGTGCAATGCGGAGAGAAAGGTGCGCTCATAGATGATGATGATGATGATGATGATGAT 1800  
DB 1748 GGTGCAATGCGGAGAGAAAGGTGCGCTCATAGATGATGATGATGATGATGATGATGAT 1807  
QY 1801 AGAGGTGCGGTGTGAGAGACAGAGTGTGCTTCCGTGTGCGGGGAGATGAGACCGAGT 1860  
DB 1808 AGAGGTGCGGTGTGAGAGACAGAGTGTGCTTCCGTGTGCGGGGAGATGAGACCGAGT 1867  
QY 1861 GGCCTTGTGAGAGAAACAGAGTGTGCGCTTGGGGGGGCCCTGAGACTTGTGATGAGAA 1920  
DB 1868 GGCCTTGTGAGAGAAACAGAGTGTGCGCTTGGGGGGGCCCTGAGACTTGTGATGAGAA 1927  
QY 1921 ATGGAGAGAAAGAGAGAGAGCTGAGAGACCTGAGAAATGAGATTAAGGCGAGACCGT 1980  
DB 1928 ATGGAGAGAAAGAGAGAGAGCTGAGAGACCTGAGAAATGAGATTAAGGCGAGACCGT 1987  
QY 1981 CAGAGAGCGCAGAGATGAGGCTTCTAGATGACAGAGACCCCGCAGA 2025  
DB 1988 CAGAGAGCGCAGAGATGAGGCTTCTAGATGACAGAGACCCCGCAGA 2032

RESULT 2  
US-09-880-107-3769  
Sequence 3769, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scheif, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02





Qy	1681	GAAGGCTTCCTCCCGCACCCCTTTCCGGCCCCGGGGTGGAGTCGGGAGAGGTGGCCCT	1740
Db	1688	GAAGGCTTCCTCCCGCACCCCTTTCCGGCCCCGGGGTGGAGTCGGGAGAGGTGGCCCT	1747
Qy	1741	GATGGCATGCGGGAGAGAAAGAGTGGCCTCATGTATCGTGTGTCTCCGCTGTAATGTTTC	1800
Db	1748	GATGGCATGCGGGAGAGAAAGAGTGGCCTCATGTATCGTGTGTCTCCGCTGTAATGTTTC	1807
Qy	1801	AGAGGTGGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGGCATGTGACCGAGGT	1866
Db	1808	AGAGGTGGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGGCATGTGACCGAGGT	1867
Qy	1861	GGCTTTGGTGGAGGAAGAAGAGTGGCCCTGTGGGGGGCCCCCTGGACCTTTGATGTGAAGAAG	1927
Db	1868	GGCTTTGGTGGAGGAAGAAGAGTGGCCCTGTGGGGGGCCCCCTGGACCTTTGATGTGAAGAAG	1927
Qy	1921	ATGGAGAGAGAAAGAGAGAGACGTGTGAAGACCTGTGAATAATGATTAAGCGAGCACCGT	1987
Db	1928	ATGGAGAGAGAAAGAGAGAGACGTGTGAAGACCTGTGAATAATGATTAAGCGAGCACCGT	1987
Qy	1961	CAGGAGCCGACAGATCGGCCCTTACTGATGACAGAGACCCCGCAGA	2025
Db	1968	CAGGAGCCGACAGATCGGCCCTTACTGATGACAGAGACCCCGCAGA	2032
RESULT 4			
US-09-873-319-717			
Sequence 717, Application US/09873319A			
Publication No. US20030134324A1			
GENERAL INFORMATION:			
APPLICANT: Munger, William E.			
APPLICANT: Kulkarni, Prakash			
APPLICANT: Getzenberg, Robert H.			
APPLICANT: Watanabe, Iwao			
APPLICANT: Yamamoto, Jun			
TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic			
TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles			
FILE REFERENCE: 44921-5029-US			
CURRENT APPLICATION NUMBER: US/09/873.319A			
CURRENT FILING DATE: 2001-06-05			
EARLIER APPLICATION NUMBER: US 60/223,323			
EARLIER FILING DATE: 2000-08-07			
NUMBER OF SEQ ID NOS: 755			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 717			
LENGTH: 2390			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899			
US-09-873-319-717			
Query Match			
Best Local Similarity 99.9%; Score 2023.4; DB 10; Length 2390;			
Matches 2024; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	GACGGACCTTGAGAGAACGAGAGAGAGAGAAATGGGCTCCACGGATTAACGTAC	60
Db	8	GACGGACCTTGAGAGAACGAGAGAGAGAGAAATGGGCTCCACGGATTAACGTAC	67
Qy	61	TATAGCCAGCTGCAGCGCCAGAGGGCTTACAGTGCTTACACCGCCAGCCCATTAAGGA	120
Db	68	TATAGCCAGCTGCAGCGCCAGAGGGCTTACAGTGCTTACACCGCCAGCCCATTAAGGA	127
Qy	121	TATGACAGACACCCAGGACATATGGGCAACAAAGCTATGGAACTTATGACAGCCCACT	180
Db	128	TATGACAGACACCCAGGACATATGGGCAACAAAGCTATGGAACTTATGACAGCCCACT	187
Qy	181	GATGTCAAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGACAGACCGCTATGTCACT	240
Db	188	GATGTCAAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGACAGACCGCTATGTCACT	247
Qy	241	TCTTATGACAGCCCTCCCACTGGTTATATCTACCTCCACCTGCCCCGAGGATTAAGCCAG	300

Db	248	TCCTATGAGCAGCTCCCACTGGTTATCTACTCACTGCCCCCAGGATACAGCCAG	307
Oy	301	CTGTCCAGGGGATATGGCACTGGTGTATATGATACCACTGCTACAGTACCAACACC	360
Db	308	CCTGTCCAGGGGATATGGCACTGGTGTATATGATACCACTGCTACAGTACCAACACC	367
Oy	361	CAGGCTTCCTATGCAAGCTCAAGTCTGCAATATGGCACTAGGCTGTTATCCAGCTTATGG	420
Db	368	CAGGCTTCCTATGCAAGCTCAAGTCTGCAATATGGCACTAGGCTGTTATCCAGCTTATGG	427
Oy	421	CAGGAGCAGAGCCATGCACTGACCTACAAAGCCGAGGATGGAAACAAGCCCACTGACACT	480
Db	428	CAGGAGCAGAGCCATGCACTGACCTACAAAGCCGAGGATGGAAACAAGCCCACTGACACT	487
Oy	481	AGTCAACCTCAATCTAGCAACAGGGGTTTACAACAGCCCAAGCTTAGATATGACAGAGT	540
Db	488	AGTCAACCTCAATCTAGCAACAGGGGTTTACAACAGCCCAAGCTTAGATATGACAGAGT	547
Oy	541	AACCTACAGTTATCCCAAGGTACTGGGAGCTACCCCATGACGCCATGCACTGCACCTTCA	600
Db	548	AACCTACAGTTATCCCAAGGTACTGGGAGCTACCCCATGACGCCATGCACTGCACCTTCA	607
Oy	601	TCCTAACCTCCTACAGCTATTCCTCTACACAGCCGACTAGTTATGATTCAGAGCAAGTTAC	660
Db	608	TCCTAACCTCCTACAGCTATTCCTCTACACAGCCGACTAGTTATGATTCAGAGCAAGTTAC	667
Oy	661	TCTCAGAGAAACACTATATGGGCAACCGAGAGCTATGAGACAGCAGATAGCTATGTCAA	720
Db	668	TCTCAGAGAAACACTATATGGGCAACCGAGAGCTATGAGACAGCAGATAGCTATGTCAA	727
Oy	721	CAAGCAGCTATATGGGCAAGACCTCCCACTAGTTATCCCACTCCCAACTGATCTCTACAGC	780
Db	728	CAAGCAGCTATATGGGCAAGACCTCCCACTAGTTATCCCACTCCCAACTGATCTCTACAGC	787
Oy	781	CAACCTCAAGTCAATATATGACCAACAAGACAGAGCTACCGGCGAGCAGACTTCAATCCGA	840
Db	788	CAACCTCAAGTCAATATATGACCAACAAGACAGAGCTACCGGCGAGCAGACTTCAATCCGA	847
Oy	841	CAGGACCAACCCAGTACATATGGGTTTATATGGGCAAGAGTCTGGAGGATTTTCCGAGCA	900
Db	848	CAGGACCAACCCAGTACATATGGGTTTATATGGGCAAGAGTCTGGAGGATTTTCCGAGCA	907
Oy	901	GGAGAGAACCCGAGCATAGATGGGCTCTGATATCCGGGGCAGGGAGAGAGGGGGATTTGAT	960
Db	908	GGAGAGAACCCGAGCATAGATGGGCTCTGATATCCGGGGCAGGGAGAGAGGGGGATTTGAT	967
Oy	961	CGTGGAGGCATATGACAGAGTGGGCGGGGAGAGAGAGCGGCTGGAATGGGCAAGCTGGA	1020
Db	968	CGTGGAGGCATATGACAGAGTGGGCGGGGAGAGAGAGCGGCTGGAATGGGCAAGCTGGA	1027
Oy	1021	GAGCGAGGTGGCTTCAATTAAGCCTGGTGGAGCCCATGGAATGAAGGACCAATCTTGTATCTA	1080
Db	1028	GAGCGAGGTGGCTTCAATTAAGCCTGGTGGAGCCCATGGAATGAAGGACCAATCTTGTATCTA	1087
Oy	1081	GGCCCACTGTAGATCCAGATGGAAGACTCTGACAAACAGTGCAAATTTATGTACAAAGATTA	1140
Db	1088	GGCCCTCTGTAGATCCAGATGGAAGACTCTGACAAACAGTGCAAATTTATGTACAAAGATTA	1147
Oy	1141	AATGACAGTGTAGCTTATGATATGATCTTGAAGCTTCTTTAAGCAGTGTGGGTTGTATAG	1200
Db	1148	AATGACAGTGTAGCTTATGATATGATCTTGAAGCTTCTTTAAGCAGTGTGGGTTGTATAG	1207
Oy	1201	ATGGAACAAGAACTGGGGCAACCATATGATCACTCACTCTGACAAAGAAACAAGGAAG	1260
Db	1208	ATGGAACAAGAACTGGGGCAACCATATGATCACTCACTCTGACAAAGAAACAAGGAAG	1267
Oy	1261	CCCAAAAGGCATATGCAAGTGTCTTATGAAAGCCCACTGCACAGGCTGCGCTGGAA	1320
Db	1268	CCCAAAAGGCATATGCAAGTGTCTTATGAAAGCCCACTGCACAGGCTGCGCTGGAA	1327
Oy	1321	TGCTTTATATGGAAGATTTTCAAGGGAGCAAACTTAAAGTCTTCCCTTCTCGAAGAG	1380

Db	1328	TGGrTTTATGGGAAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGAG	1387
Qy	1381	CCTCCAAATGAAACAGTATTCGGGGTGGTGTCCCAACCCCGTGAAGGGCAGAGGCATGCAACA	1440
Db	1388	CCTCCAAATGAAACAGTATTCGGGGTGGTGTCCCAACCCCGTGAAGGGCAGAGGCATGCAACA	1447
Qy	1441	CCACTCCGTGAGAGGTCCAGAGAGGCCCAAGAGGTCTCTGGAGGACCATGGTTCGATGGGA	1500
Db	1448	CCACTCCGTGAGAGGTCCAGAGAGGCCCAAGAGGTCTCTGGAGGACCATGGTTCGATGGGA	1507
Qy	1501	GGCGGTGAGAGGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGAAAC	1560
Db	1508	GGCGGTGAGAGGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGAAAC	1567
Qy	1561	CCCTCTGGAGAGGAAACGTCCAGAGCCGAGGTGGAGCTGGCAAGTGTCCCAATCCGGGT	1620
Db	1568	CCCTCTGGAGAGGAAACGTCCAGAGCCGAGGTGGAGCTGGCAAGTGTCCCAATCCGGGT	1627
Qy	1621	TGTGAAACCAAGAACTTGCGCTTGAGAAACAAGTGTCAACCAAGTGTAAAGGCCCAACGCT	1680
Db	1628	TGTGAAACCAAGAACTTGCGCTTGAGAAACAAGTGTCAACCAAGTGTAAAGGCCCAACGCT	1687
Qy	1681	GAAGGCTTCTTCCCGCCACCTTTTCCGGCCCGGGTGGTATTCGTGGACAGAGTGGCCCT	1744
Db	1688	GAAGGCTTCTTCCCGCCACCTTTTCCGGCCCGGGTGGTATTCGTGGACAGAGTGGCCCT	1747
Qy	1741	GATGGCATGCGGGGAGGAAGAGTGGCTCATATGATTCGTGGTGTCCCGGTGGAATGTTTC	1800
Db	1748	GATGGCATGCGGGGAGGAAGAGTGGCTCATATGATTCGTGGTGTCCCGGTGGAATGTTTC	1807
Qy	1801	AGAGGTGGCGCTGTGTGGAGACAGAGGTGGCTTCGTTGGTGGCCGAGGACATGGACCGAGGT	1866
Db	1808	AGAGGTGGCGCTGTGTGGAGACAGAGGTGGCTTCGTTGGTGGCCGAGGACATGGACCGAGGT	1867
Qy	1861	GGCTTTGGTGGAGGAAGACGAGAGTGGCCCTGGGGGGGCCCTTGGAACCTTTGATGGACAG	1920
Db	1868	GGCTTTGGTGGAGGAAGACGAGAGTGGCCCTGGGGGGGCCCTTGGAACCTTTGATGGACAG	1922
Qy	1921	ATGGAGGAAGAAAGAGAGGAGCTGTGAGAGACTTGGAAAAATGGATTAAGCGAGCACCGT	1980
Db	1928	ATGGAGGAAGAAAGAGAGGAGCTGTGAGAGACTTGGAAAAATGGATTAAGCGAGCACCGT	1987
Qy	1981	CAGAGCGCCAGAGATCGGCCCTCACTAAGATGCAAGAACCCCGGAGA	2025
Db	1988	CAGAGCGCCAGAGATCGGCCCTCACTAAGATGCAAGAACCCCGGAGA	2032

```

/ RESULT 5
/ US-09-822-830A-49/c
/ Sequence 49, Application US/09822830A
/ Patent No. US20020142952A1
/ GENERAL INFORMATION:
/ APPLICANT: Genetics Institute, Inc.
/ APPLICANT: Wong, Gordon G.
/ APPLICANT: Clark, Hilary
/ APPLICANT: Fechtel, Kim
/ APPLICANT: Agostino, Michael J.
/ APPLICANT: Howes, Steven H.
/ APPLICANT: Reenick, Richard J.
/ APPLICANT: Gulukota, Kamalakara
/ APPLICANT: Graham, James R.
/ TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL, SECRETED PROTEINS
/ FILE REFERENCE: GIN 6402
/ CURRENT APPLICATION NUMBER: US/09/822,830A
/ CURRENT FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/195,604
/ PRIOR FILING DATE: 2000-04-06
/ NUMBER OF SEQ ID NOS: 631
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 49
/ LENGTH: 2176
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

Query Match	Best Local Similarity	98.1% Matches 1998;	Score 1986.4;	DB 9;	Length 2176;	
		0;	Mismatches 1;	Indels 1;	Gaps 1;	
QY	26	AMGAGAGAAAATGGCTCCAGGATTAACAGTACCTATTAGCCAAAGCTGCAGCCAGCAGG	85			
DB	2171	AMGTTGAGAAAATGGCTCCAGGATTAACAGTACCTATTAGCCAAAGCTGCAGCCAGCAGG	2112			
QY	86	GCTACAGTGTCTTACACCGCCAGCCCACTCAAGGATATGCAACAGACCACCGCATATG	145			
DB	2111	GCTACAGTGTCTTACACCGCCAGCCCACTCAAGGATATGCAACAGACCACCGCATATG	2052			
QY	146	GGCAACAAAGTATGGAACCTATGGAACAGCCCACTGATATGACGTTATACCAAGGCTCAGA	205			
DB	2051	GGCAACAAAGTATGGAACCTATGGAACAGCCCACTGATATGACGTTATACCAAGGCTCAGA	1992			
QY	206	CCACTGCAACCTATGGGAGACCGCCCTATGCAACTTTTATGGAACAGCTTCCACTGTGTT	265			
DB	1991	CCACTGCAACCTATGGGAGACCGCCCTATGCAACTTTTATGGAACAGCTTCCACTGTGTT	1932			
QY	266	ATACTACTCCAACTGCCCCCAAGGCATACAGCCCTGTCCAGGGTATGGACTGTGTG	325			
DB	1991	ATACTACTCCAACTGCCCCCAAGGCATACAGCCCTGTCCAGGGTATGGACTGTGTG	1872			
QY	326	CTTATGATTAACCAACTGTCTACAGTACACCAACCCAGGCTCTCTATGACGTCTACGTCTG	385			
DB	1871	CTTATGATTAACCAACTGTCTACAGTACACCAACCCAGGCTCTCTATGACGTCTACGTCTG	1812			
QY	386	CATATGGCACTCAGCCTGTATATCAGCTATATGGGGAGAGCCAGCCAGCTACCTTA	445			
DB	1811	CATATGGCACTCAGCCTGTATATCAGCTATATGGGGAGAGCCAGCCAGCTACCTTA	1752			
QY	446	CAAGACGCGCAGAGATGGAACAAGCCCACTGATCTAAGTCAACTCAATCTTAGCAACAGGG	505			
DB	1751	CAAGACGCGCAGAGATGGAACAAGCCCACTGATCTAAGTCAACTCAATCTTAGCAACAGGG	1693			
QY	506	GTTACAAACAGCCAGCTTATGATATGGAACAGTAACCTACAGTTATCCCAAGTACCTG	565			
DB	1692	GTTACAAACAGCCAGCTTATGATATGGAACAGTAACCTACAGTTATCCCAAGTACCTG	1633			
QY	566	GGACCTACCCCATGACAGCAGTCACTGCACTCATCTCACTCCCTCCACAGCAATTCCT	625			
DB	1632	GGACCTACCCCATGACAGCAGTCACTGCACTCATCTCACTCCCTCCACAGCAATTCCT	1573			
QY	626	CTACACAGCCGACATGATTATGATCAGAGCAGTTACTCTCAGCAGAACCACTATGGGCAAC	685			
DB	1572	CTACACAGCCGACATGATTATGATCAGAGCAGTTACTCTCAGCAGAACCACTATGGGCAAC	1513			
QY	686	CGAGCAGCTATGACAGCAGAGTAGCTATGCTCAACAAAGCATATGGGAGCAGACCTCT	745			
DB	1512	CGAGCAGCTATGACAGCAGAGTAGCTATGCTCAACAAAGCATATGGGAGCAGACCTCT	1453			
QY	746	CCACTAGTACCCACCCCAACCTGAGTCTTACAGCCAAAGCTCCAAAGTCAATATATGCCAAC	805			
DB	1452	CCACTAGTACCCACCCCAACCTGAGTCTTACAGCCAAAGCTCCAAAGTCAATATATGCCAAC	1393			
QY	806	AGAGCAGCAGCTACGGGAGCAGAGTTCATTTCCAGACAGAACCAACCCAGTATGAGTGTG	865			
DB	1392	AGAGCAGCAGCTACGGGAGCAGAGTTCATTTCCAGACAGAACCAACCCAGTATGAGTGTG	1333			
QY	866	TTTATGGCAGAGATCTTGAGGATTTTTCGGGACAGAGAGAACCGGAGCATGAGTGTG	925			
DB	1332	TTTATGGCAGAGATCTTGAGGATTTTTCGGGACAGAGAGAACCGGAGCATGAGTGTG	1273			
QY	926	CTGATTAACCGGGGCGGGGAAAGAGGGGATTTGATCTGTGAGGCAATGACGAGAGTGGGC	985			
DB	1272	CTGATTAACCGGGGCGGGGAAAGAGGGGATTTGATCTGTGAGGCAATGACGAGAGTGGGC	1213			
QY	986	GGGAGAGAGAGCGGGGTGGAATGGGAGACGCTGAGAGCGGAGGTGGCTTCAATTAAGCTGT	1045			
DB	1212	GGGAGAGAGAGCGGGGTGGAATGGGAGACGCTGAGAGCGGAGGTGGCTTCAATTAAGCTGT	1153			

OY	1046	GTGAGCCCATGGATGAAGAACAAGACTCTTGATCTTAGGGCCACCTGTATATCCAGATGAG	1105
Db	1152	GTGAGCCCATGGATGAAGAACAAGACTCTTGATCTTAGGGCCACCTGTATATCCAGATGAG	1093
OY	1106	ACTCTGAACAAGTGCAAATTTATATGACAGATTTAAATGACAGTGTGACTCTAGATGATC	1165
Db	1092	ACTCTGAACAAGTGCAAATTTATATGACAGATTTAAATGACAGTGTGACTCTAGATGATC	1033
OY	1166	TGGCAGACTTCTTTAAGCAGTGTGGGGTTTGTTAAGATGACAGAGAACTGGGCAACCA	1225
Db	1032	TGGCAGACTTCTTTAAGCAGTGTGGGGTTTGTTAAGATGACAGAGAACTGGGCAACCA	973
OY	1226	TGATCCACATCTACCTGGACAAAGAAACAGGAAACCCAAAGGCGCATGCCACAGTGTCT	1285
Db	972	TGATCCACATCTACCTGGACAAAGAAACAGGAAACCCAAAGGCGCATGCCACAGTGTCT	913
OY	1286	ATGAAGAACCACCACTGCACAGGCTGCCTGTGAAATGGTTTGAATGGAGAAATTTCAAG	1345
Db	912	ATGAAGAACCACCACTGCACAGGCTGCCTGTGAAATGGTTTGAATGGAGAAATTTTCAAG	853
OY	1346	GGAGCAAACTTAAAGTCTTCCTTGTCTCGGAGAGACCTTCATATGAACTGTATCGGGGTG	1405
Db	852	GGAGCAAACTTAAAGTCTTCCTTGTCTCGGAGAGACCTTCATATGAACTGTATCGGGGTG	793
OY	1406	GTCTGCCACCCCGTGAAGGGCAGAGGATGGCCACCACTCCCGTGAAGGTCCAGAGGCTC	1465
Db	792	GTCTGCCACCCCGTGAAGGGCAGAGGATGGCCACCACTCCCGTGAAGGTCCAGAGGCTC	733
OY	1466	CAGAGAGTCTTGGGGGAGCCCAATGGGTGCAATGGAGAGCGCTGTGAAGATAGAGAGCT	1525
Db	732	CAGAGAGTCTTGGGGGAGCCCAATGGGTGCGCATGGAGAGCGCTGTGAAGATAGAGAGCT	673
OY	1526	TCCCTCCCAAGAGAACCCCGGGGTTTCCCGAGGGAACCCCTCTTGAGAGAGAAACGTCCAGC	1585
Db	672	TCCCTCCCAAGAGAACCCCGGGGTTTCCCGAGGGAACCCCTCTTGAGAGAGAAACGTCCAGC	613
OY	1586	ACCGAGCTGAGAGACTGCGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTCGCTGGA	1645
Db	612	ACCGAGCTGAGAGACTGCGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTCGCTGGA	553
OY	1646	GAAACAGATGCACAAGTGTAAAGGCCCAAAAGCTGAAGGCTTCTCCCGCACCCCTTTC	1705
Db	552	GAAACAGATGCACAAGTGTAAAGGCCCAAAAGCTGAAGGCTTCTCTCCCGCACCCCTTTC	493
OY	1706	CGCCCCCGGGTGTGTATCTGTGGCAAGGTGGCCCTGTGTGCATGTGGGGAAGAAAGGTG	1765
Db	492	CGCCCCCGGGTGTGTATCTGTGGCAAGGTGGCCCTGTGTGCATGTGGGGAAGAAAGGTG	433
OY	1766	GCTTCATAGATCGTGTGTGTGTCCCGGTGGAATGTTTCAAGAGTGTGCGGTGTGTGAAGACAG	1825
Db	432	GCTTCATAGATCGTGTGTGTGTGTCCCGGTGGAATGTTTCAAGAGTGTGCGGTGTGTGAAGACAG	373
OY	1826	GTGGCTTCCGTGTGTGCGCGGGCAATGACCCGAGGTGTGCTTTTGTGTGAAGAAACGAGGTG	1885
Db	372	GTGGCTTCCGTGTGTGCGCGGGCAATGACCCGAGGTGTGCTTTTGTGTGAAGAAACGAGGTG	313
OY	1886	GCCCTGGGGGGCCCCCTTGAGACTTTTGAATGGAACAGATGGAGAGAAAGAGAGAGAGCTG	1945
Db	312	GCCCTGGGGGGCCCCCTTGAGACTTTTGAATGGAACAGATGGAGAGAAAGAGAGAGAGCTG	253
OY	1946	GAGGACTGTGAAAAATGTATAAGGCGAGACCCGTCAAGAGCGCAGAGATCGGCGCTTACT	2005
Db	252	GAGGACTGTGAAAAATGTATAAGGCGAGACCCGTCAAGAGCGCAGAGATCGGCGCTTACT	193
OY	2006	AGATGCAAGAACCCCGCAGA 2025	
Db	192	AGATGCAAGAACCCCGCAGA 173	

```

Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Reenick, Richard J.
APPLICANT: Gulboka, Kamalakar
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 410
LENGTH: 2273
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-410

```

Query Match		91.7%;	Score 1856.6;	DB 9;	Length 2273;
Best Local Similarity		97.1%;	Pred. No. 0;		
Matches 1933;		Conservative	0;	Mismatches	4;
				Indels	54;
				Gaps	2;
Qy	35	AAATGGCTTCACGGATTTACAGTACCTTATAGCCAGCTGCAGCGCAGACAGGGCTTACAGTG	94		
Db	2268	AAAGGTGGGTCCAGCGATTTACAGTACCTTATAGCCAGCTGCAGCGCAGGGCTTACAGTG	2209		
Qy	95	CTTTCACCGCCAGCCCACTCAAGATATGACAAGACACCCAGGCAATATGGCAACAA	154		
Db	2208	CTTTCACCGCCAGCCCACTCAAGATATGACAAGACACCCAGGCAATATGGCAACAA	2149		
Qy	155	GCTATGGAACCTTATGACAGCCCACTGATGTCACTTATCCAGGCTCAGACCACTGCAA	214		
Db	2148	GCTATGGAACCTTATGACAGCCCACTGATGTCACTTATCCAGGCTCAGACCACTGCAA	2089		
Qy	215	CCATATGGGACGACCGCCCTATGCAACTTCTTATAGACAGCCTCCCACTGGTATATCTATC	274		
Db	2088	CCATATGGGACGACCGCCCTTATGCAACTTCTTATAGACAGCCTCCCACTGGTATATCTATC	2029		
Qy	275	CAACTGGCCCCCAGGACATACAGCAGCCTGTCCAGGGGTATGGACTGTGCTTATGATA	334		
Db	2028	CAACTGGCCCCCAGGACATACAGCAGCCTGTCCAGGGGTATGGACTGTGCTTATGATA	1969		
Qy	335	CCACCACTGCTTACAGTACCAACAACCAAGGCTCTCTATGCAAGTCAAGTTCGATATGGCA	394		
Db	1968	CCACCACTGCTTACAGTACCAACAACCAAGGCTCTCTATGCAAGTCAAGTTCGATATGGCA	1909		
Qy	395	CTCAGCCTGTCTTATCCAGCCTATATGGGACAGGACAGGACACTGCACCTTACAAGACCGC	454		
Db	1908	CTCAGCCTGTCTTATCCAGCCTATATGGGACAGGACAGGACACTGCACCTTACAAGACCGC	1849		
Qy	455	AGGATGGAAAACAAGCCCACTGAGACTATGTCAACCTCAATTTAGACAGGGGGTTACAAC	514		
Db	1848	AGGATGGAAAACAAGCCCACTGAGACTATGTCAACCTCAATTTAGACAGGGGGTTACAAC	1789		
Qy	515	AGCCCAAGCCTTATATGACAAGATATCAACAGTATATCCCAAGGTACTGGGACGTAC	574		
Db	1788	AGCCCAAGCCTTATATGACAAGATATCAACAGTATATCCCAAGGTACTGGGACGTAC	1729		
Qy	575	CCATGCAAGCAGTCACTGCACCTTCATCCTTACCTCTTACAGTATTTCTTACACAGC	634		
Db	1728	CCATGCAAGCAGTCACTGCACCTTCATCCTTACCTCTTACAGTATTTCTTACACAGC	1669		
Qy	635	CGACTATGTTATATCAAGAGATTACTCTACAGAGAAACAATATGGGCAACCGACAGCT	694		
Db	1668	CGACTATGTTATATCAAGAGATTACTCTACAGAGAAACAATATGGGCAACCGACAGCT	1609		



QY 695 ATGAGCAGAGATAGTATGTCACAAAGCAGTATGAGGCGACGCTCCCACTAGTT 754  
DB 1608 ATGAGCAGAGATAGTATGTCACAAAGCAGTATGAGGCGACGCTCCCACTAGTT 1549  
QY 755 ACCACCCCAAGCTGATCTTACAGCCAGCTCCAGTCAATATAGCCAAACAGAGCAGA 814  
DB 1548 ACCACCCCAAGCTGATCTTACAGCCAGCTCCAGTCAATATAGCCAAACAGAGCAGA 1489  
QY 815 GCTACGGGCGACGAGGTTCAATCCGACAGACCAACCCCACTAGTATGAGTGTATAGGCG 874  
DB 1488 GCTACGGGCGACGAGGTTCAATCCGACAGACCAACCCCACTAGTATGAGTGTATAGGCG 1429  
QY 875 AGGAGTGTAGAGATTTTCCGACACAGAGAGAACCGGAGATAGAGTGGCCCTGATAGC 934  
DB 1428 AGGAGTGTAGAGATTTTCCGACACAGAGAGAACCGGAGATAGAGTGGCCCTGATAGC 1369  
QY 935 GGGGCGAGGAGAGAGGAGGATTTGATGCTGAGAGCAGAGTGGCGGAGAGAG 994  
DB 1368 GGGGCGAGGAGAGAGGAGGATTTGATGCTGAGAGCAGAGTGGCGGAGAGAG 1309  
QY 995 GAGCGGCTGAGATGCGGAGCGCTGAGAGCAGAGTGGCTTCAATAGCTGTGAGACCA 1054  
DB 1308 GAGCGGCTGAGATGCGGAGCGCTGAGAGCAGAGTGGCTTCAATAGCTGTGAGACCA 1252  
QY 1055 TGAGTGAAGAGCAGATCTTATAGTATAGGCGCACCTGTAGATCCAGATGAAGACTGACA 1114  
DB 1251 TGAGTGAAGAGCAGATCTTATAGTATAGGCGCACCTGTAGATCCAGATGAAGACTGACA 1192  
QY 1115 ACAAGTCAATTTATGTAACAAGATTAATGACAGTGTGACTAGATCTGAGCAGACT 1174  
DB 1191 ACAAGTCAATTTATGTAACAAGATTAATGACAGTGTGACTAGATCTGAGCAGACT 1132  
QY 1175 TCTTTAAGCAGATGCGGCTGTGTAAGTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 1234  
DB 1131 TCTTTAAGCAGATGCGGCTGTGTAAGTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 1072  
QY 1235 TCTACCTGAGCAAG 1294  
DB 1071 TCTACCTGAGCAAG 1012  
QY 1295 CACCCACTGCGCAAGGCTGCGCTGAGATGTTGATGAGAGAGATTTTCAAGAGAGAGAG 1354  
DB 1011 CACCCACTGCGCAAGGCTGCGCTGAGATGTTGATGAGAGAGATTTTCAAGAGAGAGAG 952  
QY 1355 TTTAAAGTCTCTCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1414  
DB 951 TTTAAAGTCTCTCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 892  
QY 1415 CCGGTGAG 1474  
DB 891 CCGGTGAG 832  
QY 1475 CTGGGAG 1534  
DB 831 CTGGGAG 772  
QY 1535 GAG 1594  
DB 771 GAG 712  
QY 1595 GAG 1654  
DB 711 GAG 652  
QY 1655 GCAACCAAGTGAAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1714  
DB 651 GCAACCAAGT----- 643  
QY 1715 GTGGTGAATGCTGCGAG 1774  
DB 642 GTGGTGAATGCTGCGAG 583  
QY 1775 ATCGGTGAGTCCGGGTGAGATGTTCAAGAGTGGCGGTGAGAGAGAGAGAGAGAGAGTGC 1834

DB 582 ATCGGTGAGTCCGGGTGAGATGTTCAAGAGTGGCGGTGAGAGAGAGAGAGAGAGTGC 523  
QY 1835 GTGGTGGCGGAG 1894  
DB 522 GTGGTGGCGGAG 463  
QY 1895 GGGCGGCTGAG 1954  
DB 462 GGGCGGCTGAG 403  
QY 1955 GAAAAATGATTAAGGCGAG 2014  
DB 402 GAAAAATGATTAAGGCGAG 343  
QY 2015 GACCCCGCAGAG 2025  
DB 342 GACCCCGCAGAG 332

RESULT 7  
US-10-094-749-1074  
; Sequence 1074, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YUKI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKI, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 08435/0160  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1074  
; LENGTH: 1988  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-1074  
Query Match 81.6%; Score 1652.4; DB 17; Length 1988;  
Best Local Similarity 91.5%; Pred. No. 0;  
Matches 1831; Conservative 0; Mismatches 1; Indels 168; Gaps 1;  
QY 26 AAG 85  
DB 1 AAG 60  
QY 86 GCTACAGTGTCTTACCGCCAG 145  
DB 61 GCTACAGTGTCTTACCGCCAG 120  
QY 146 GCGAACAAGCTATGAACTATGACAGCCCACTGATGCTAGCTATATACCAGGCTCAGA 205

Db 121 GGCAACAAAGCTATGAACTTATGAGAGCCCACTGATGCTATATCCAGGCTCAGA 180  
Qy 206 CCACTGCAACCTATGAGGAGAGCCGCTATGCAATTTCTTATGAGACGCTCCACTG3TT 265  
Db 181 CCACTGCAACCTATGAGGAGAGCCGCTATGCAATTTCTTATGAGACGCTCCACTG3TT 240  
Qy 266 ATACTACTCCAACTGCCCCCGAGGATACAGCCAGGCTGTCAGGGGTAATGGGCACTG3TG 325  
Db 241 ATACTACTCCAACTGCCCCCGAGGATACAGCCAGGCTGTCAGGGGTAATGGGCACTG3TG 300  
Qy 326 CTATATGATACCACTGCTATCACTGCAACCAACCCAGGCTCTCTATGCACTGCTGCTG 385  
Db 301 CTATATGATACCACTGCTATCACTGCAACCAACCCAGGCTCTCTATGCACTGCTGCTG 360  
Qy 386 CATATGCACTCAAGCTGCTTATATCAAGCTTATGAGGAGAGCCAGGAGCCACTGCACTTA 445  
Db 361 CATATGCACTCAAGCTGCTTATATCAAGCTTATGAGGAGAGCCAGGAGCCACTGCACTTA 420  
Qy 446 CAAGACCGCAAGATGAAACAAAGCCCACTGAGACTATCAACCTCATCTAGCAAGAGGG 505  
Db 421 CA----- 422  
Qy 506 GTTACAAACAGCCAGCCTAGATATGAGACAGAGTAATCAAGTTATCCCAAGTACCTG 565  
Db 423 ----- 422  
Qy 566 GGAAGCTACCCCACTGAGCCAGCTCACTGCACTCTCATCTTACCTCTCAACGCTATTTCT 625  
Db 423 -----AGCTATTTCT 432  
Qy 626 CTACACAGCCGACTAGTTATGATGAGAGCACTTACTGAGAGAAACCTATGAGGAGAAC 685  
Db 433 CTACACAGCCGACTAGTTATGATGAGAGCACTTACTGAGAGAAACCTATGAGGAGAAC 492  
Qy 686 CGAGCACTATGAGACAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATG 745  
Db 493 CGAGCACTATGAGACAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATG 552  
Qy 746 CCACTAGTTACCCCACTGAGATGATGATGATGATGATGATGATGATGATGATGATG 805  
Db 553 CCACTAGTTACCCCACTGAGATGATGATGATGATGATGATGATGATGATGATGATG 612  
Qy 806 AGAGCAGCAGCTACGAGGAGAGAGTTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 865  
Db 613 AGAGCAGCAGCTACGAGGAGAGAGTTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 672  
Qy 866 TTATATGAGCAGAGTCTGAGAGATTTTCCGAGACAGAGAGAGAGAGAGAGAGAGAG 925  
Db 673 TTATATGAGCAGAGTCTGAGAGATTTTCCGAGACAGAGAGAGAGAGAGAGAGAGAG 732  
Qy 926 CTGATTAACCGGGGAG 985  
Db 733 CTGATTAACCGGGGAG 792  
Qy 986 GGGGAG 1045  
Db 793 GGGGAG 852  
Qy 1046 GTGAGACCATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105  
Db 853 GTGAGACCATGATGATGAG 912  
Qy 1106 ACTCTGCAACAGTGCATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1165  
Db 913 ACTCTGCAACAGTGCATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 972  
Qy 1166 TGGCAGACTCTTTAAGCAGATGAGGAGTTTAAAGAGAGAGAGAGAGAGAGAGAGAG 1225  
Db 973 TGGCAGACTCTTTAAGCAGATGAGGAGTTTAAAGAGAGAGAGAGAGAGAGAGAGAG 1032  
Qy 1226 TGATCCACATCTACTGAG 1285  
Db 1033 TGATCCACATCTACTGAG 1092

Qy 1286 ATGAGAGCCCAACCCACTGCGCAAGGCTGCGGTGATGATGATGATGATGATGATGATGAT 1345  
Db 1093 ATGAGAGCCCAACCCACTGCGCAAGGCTGCGGTGATGATGATGATGATGATGATGATGAT 1152  
Qy 1346 GAGAGCAACTTAAAGTCTCCCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1405  
Db 1153 GAGAGCAACTTAAAGTCTCCCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1212  
Qy 1406 GTCTGCCAACCCTGAG 1465  
Db 1213 GTCTGCCAACCCTGAG 1272  
Qy 1466 CAGAGAGTCTGAG 1525  
Db 1273 CAGAGAGTCTGAG 1332  
Qy 1526 TCCCTCCAG 1585  
Db 1333 TCCCTCCAG 1392  
Qy 1586 ACCGAGCTGAG 1645  
Db 1393 ACCGAGCTGAG 1452  
Qy 1646 GAAACAGAGTCAACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705  
Db 1453 GAAACAGAGTCAACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1512  
Qy 1706 GCGCCCGAG 1765  
Db 1513 GCGCCCGAG 1572  
Qy 1766 GCCTCATGATCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1825  
Db 1573 GCCTCATGATCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1632  
Qy 1826 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1885  
Db 1633 GTGGCTTCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1692  
Qy 1886 GCCCTGG 1945  
Db 1693 GCCCTGG 1752  
Qy 1946 GAGAGCCTGAG 2005  
Db 1753 GAGAGCCTGAG 1812  
Qy 2006 AGATGCAAG 2055  
Db 1813 AGATGCAAG 1832

RESULT 8  
US-10-425-115-182496  
; Sequence 182496, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; NUMBER OF SEQ ID NOS: 2003-04-28  
; SEQ ID NO 182496  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Zea mays

FEATURE:  
NAME/KEY: unsure  
LOCATION: (1..763)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_98018C.1  
US-10-425-115-182496

Query Match 31.2%: Score 631, DB 18; Length 763;  
Best Local Similarity 92.8%: Pred. No. 9.3e-174;  
Matches 661; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 29 GAGAGAAATGGGTCACGAGATTAAGTACCTATAGCCAAAGCTGACGCGACAGAGGCT 88  
DB 52 GCGAGAAATGGGTCACGAGATTAAGTACCTATAGCCAAAGCTGACGCGACAGAGGCT 111  
QY 89 ACAGTGTTCACCGCCGACCCCACTCAAGAGATATGACAGACCCAGCATATGGGC 148  
DB 112 ACAGTGTTCACCGCCGACCCCACTCAAGAGATATGACAGACCCAGCATATGGGC 171  
QY 149 AACAAAGCTATGGAACCTATGAGACGCCCTGATGCTACCTATACCCAGGCTCAGACA 208  
DB 172 AACAAAGCTATGGAACCTATGAGACGCCCTGATGCTACCTATACCCAGGCTCAGACA 231  
QY 209 CTGCAACTATGGGACAGCGCCCTATGCACTTCTTATGAGACAGCCCTCCAGCTGTTATA 268  
DB 232 CTGCAACTATGGGACAGCGCCCTATGCACTTCTTATGAGACAGCCCTCCAGCTGTTATA 291  
QY 269 CTACTCCAACTGCCCCCGACGATACAGCCAGCCCTGTCAGGGGTATGGCATGTCCTT 328  
DB 292 GCACTCCAACTGCCCCCGACGATACAGCCAGCCCTGTCAGGGGTATGGCATGTCCTT 351  
QY 329 ATGATACCACTGCTGCTACAGTCAACCAACCCAGGCTCTTATGAGGCTCAGTTCAT 388  
DB 352 ATGATACCACTGCTGCTACAGTCAACCAACCCAGGCTCTTATGAGGCTCAGTTCAT 411  
QY 389 ATGCACTGACGCTGCTTATTCAGCCTATGAGGCGACGACGACCACTGCACTCAAA 448  
DB 412 ATGCACTGACGCTGCTTATTCAGCCTATGAGGCGACGACGACCACTGCACTCAAA 471  
QY 449 GACCGCAGATGGAACCAACCCCACTGAGACTAGTCAACTCAATCTTATGACAGGGGCT 508  
DB 472 GACCGCAGATGGAACCAACCCCACTGAGACTAGTCAACTCAATCTTATGACAGGGGCT 531  
QY 509 ACAACCGCCGACGCTGATGATGAGACAGATTAAGTATATCCCAAGTACCTGGA 568  
DB 532 ATTAACCAACCCGACGCTGATGATGAGACAGATTAAGTATATCCCAAGTACCTGGA 591  
QY 569 GCTACCCGATGACGACGATGACCTGACCTCCTACCTCCTTATCCAGGCTATTCCTTA 628  
DB 592 GCTACCCGATGACGACGATGACCTGACCTCCTACCTCCTTATCCAGGCTATTCCTTA 651  
QY 629 CACAGCCGACTGATGATGATGAGACAGATTAAGTATGAGACCACTTATGGGCAACGA 688  
DB 652 CACAGCCGACTGATGATGATGAGACAGATTAAGTATGAGACCACTTATGGGCAACGA 711  
QY 689 GCAGCTATGACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 740  
DB 712 GCAGCTATGACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 763

## RESULT 9

US-10-198-846-9847  
Sequence 9847, Application US/1019846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Xu, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF BREAST CANCER

FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9847  
LENGTH: 2299  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-198-846-9847

Query Match 28.1%: Score 570; DB 14; Length 2299;  
Best Local Similarity 100.0%: Pred. No. 9.9e-156;  
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1048 GAGCCATGATGAGAGACAGATCTTATGATCTAGGCCCACTGTATCCAGATGAAGAC 1107  
DB 1072 GAGCCATGATGAGAGACAGATCTTATGATCTAGGCCCACTGTATCCAGATGAAGAC 1131  
QY 1108 TCTGACACAGTGCATTTATGATCAAGATTAATGACAGTGTGCTTAAGTATCTG 1167  
DB 1132 TCTGACACAGTGCATTTATGATCAAGATTAATGACAGTGTGCTTAAGTATCTG 1191  
QY 1168 GCAAGCTTCTTAAAGAGTGGGGGTGTTAAGATGACAGAGAACTGGGCAACCATG 1227  
DB 1192 GCAAGCTTCTTAAAGAGTGGGGGTGTTAAGATGACAGAGAACTGGGCAACCATG 1251  
QY 1228 ATCCATCTTACCTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCAGTGTCTAT 1287  
DB 1252 ATCCATCTTACCTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCAGTGTCTAT 1311  
QY 1288 GAAGACCCACCACTGCGCAAGGCTGCGGTGGAATGTTGATGGGAAAGTTTCAAGGG 1347  
DB 1312 GAAGACCCACCACTGCGCAAGGCTGCGGTGGAATGTTGATGGGAAAGTTTCAAGGG 1371  
QY 1348 AGCAACTTAAAGTCTCCCTGCTGGAGAGAAAGCCCTCAATGAAAGTATGCGGGTGT 1407  
DB 1372 AGCAACTTAAAGTCTCCCTGCTGGAGAGAAAGCCCTCAATGAAAGTATGCGGGTGT 1431  
QY 1408 CTGCGACCCCGTGAAGGAGAGAGCATGCCACCACTCCTGAGAGTCCAGAGGCCCA 1467  
DB 1432 CTGCGACCCCGTGAAGGAGAGAGCATGCCACCACTCCTGAGAGTCCAGAGGCCCA 1491  
QY 1468 GAGGTCTCTGGGAGACCATGAGTGGTCCATGAGAGGCGGTGAGAGATGAGAGGCTTC 1527  
DB 1492 GAGGTCTCTGGGAGACCATGAGTGGTCCATGAGAGGCGGTGAGAGATGAGAGGCTTC 1551  
QY 1528 CTTCAAGAGAGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGGAAAGCTCCAGAC 1587  
DB 1552 CTTCAAGAGAGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGGAAAGCTCCAGAC 1611  
QY 1588 CGAGCTGAGAGCTGAGAGTGTCCCAATCCG 1617  
DB 1612 CGAGCTGAGAGCTGAGAGTGTCCCAATCCG 1641

## RESULT 10

US-09-864-761-8001  
Sequence 8001, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312

```

RESULT 11
US-09-918-995-27690
; Sequence 27690, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27690
; LENGTH: 550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(550)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27690

```

Query Match	25.1%	Score 508.4	DB 10	Length 550
Best Local Similarity	97.5%	Pred. No. 66-138		
Matches 515	Conservative	0	Mismatches 13	Indels 0
			Gaps	0
QY	1327	GATGGGAAGAATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTGGAGAAAGCCTCCA	138	
Db	22	GAGAGGGCAATTCCTTTCAGGGAGCAAACTTAAGTCTCCCTTGCTGGAGAAAGCCTCCA	81	
QY	1387	ATGAACAGTATGCGGGGTGCTTCGCCACCCTGAGGGCAGAGGCAATGCCACCACTC	144	
Db	82	ATGAACAGTATGCGGGGTGCTTCGCCACCCTGAGGGCAGAGGCAATGCCACCACTC	141	
QY	1447	CGTGAAGTCCAGAGGGCCCAAGAGTCTTGGGGGACCATGGGTGCATGGAGGCGGT	150	
Db	142	CGTGAAGTCCAGAGGGCCCAAGAGTCTTGGGGGACCATGGGTGCATGGAGGCGGT	201	
QY	1507	GGAGGAGATGAGAGGAGGCTCCCTCCAGAGGACCCCGGGGTTCCCGAGAGGAACCCCTCT	156	
Db	202	GGAGGAGATGAGAGGAGGCTCCCTCCAGAGGACCCCGGGGTTCCCGAGAGGAACCCCTCT	261	
QY	1567	GGAGGAGAAACGTCCAGCACCCGAGCTGGAGACTG3CAGTGTCCCAATCCGGGTTTGGGA	162	
Db	262	GGAGGAGAAACGTCCAGCACCCGAGCTGGAGACTGGCACTGTCCCAATCCGGGTTTGGGA	321	
QY	1627	AACCAAACTTGCGCTGGAGAAACAGAGTGCAACCAAGTTAAG3CCCAAAGCCTTAGAGC	168	
Db	322	AACCAAACTTGCGCTGGAGAAACAGAGTGCAACCAAGTTAAG3CCCAAAGCCTTAGAGC	381	

QY 1687 TTCTCCCGCCACCCCTTCCGCCCGGGTGTGATCTGTGCGCAGAGGTGGCCCTGGTGGC 1746  
DB 382 TTCTCCCGCCACCCCTTCCGCCCGGGTGTGATCTGTGCGCAGAGGTGGCCCTGGTGGC 441  
QY 1747 ATCGGGGAGAGAGAGGTGGCTCATGGATCGTGTGTCTCCGGTGGATGTTCAAGGT 1806  
DB 442 ATCGGGGAGAGAGAGGTGGCTCATGGATCGTGTGTCTCCGGTGGATGTTCAAGGT 501  
QY 1807 GGGCGGTGTGAGAGACAGAGGTGGCTTCCGTGTGGCCGGGAGCATGAGC 1854  
DB 502 GGGCGGTGTGAGAGACAGAGGTGGCTTCCGTGTGGCCGGGAGCATGAGC 549  
RESULT 12  
US-09-918-995-9557  
Sequence 9557, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9557  
LENGTH: 568  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) ... (568)  
OTHER INFORMATION: n = A, T, C or G  
US-09-918-995-9557  
Query Match 23.7%; Score 480.2; DB 10; Length 568;  
Best Local Similarity 95.5%; Pred. No. 1,1e-129;  
Matches 515; Conservative 0; Mismatches 20; Indels 4; Gaps 2;  
QY 932 ACCGGGCGAGGAGAGAGGAGGAGGATTTGATCTGTGAGGATGAGAGAGGTGGGGGAG 991  
DB 33 ATCGGTGTGTGATTCGAGATTGATCTGTGAGGATGAGAGAGGTGGGGGAG 92  
QY 992 GAGGACGGGTGTGATGTGGGAGCGCTGTGAGAGGAGGTGGCTTCAATAGCCTGTGAGC 1051  
DB 93 GAGGACGGGTGTGATGTGGGAGCGCTGTGAGAGGAGGTGGCTTCAATAGCCTGTGAGC 149  
QY 1052 CCATGATGAGAGACAGATCTTGTATGAGCCCACTGTAGATCCAGATGAGACTGTG 1111  
DB 150 CCATGATGAGAGACAGATCTTGTATGAGCCCACTGTAGATCCAGATGAGACTGTG 209  
QY 1112 ACAACAGTGAATTTATGATCAAGATTAATGACGTGTGACCTGTAGATGATCTGGCAG 1171  
DB 210 ACAACAGTGAATTTATGATCAAGATTAATGACGTGTGACCTGTAGATGATCTGGCAG 269  
QY 1172 ACTCTTTAAGACGTGTGGGTTGTTAAGATGAGCAAGAGAACTGGGCAACCATGATCC 1231  
DB 270 ACTCTTTAAGACGTGTGGGTTGTTAAGATGAGCAAGAGAACTGGGCAACCATGATCC 329  
QY 1232 ACATCTACTGTGAGCAAGAGAAACAGAGAAAGCCCAAGGGAGTGCACAGTGTCTATGAG 1291  
DB 330 ACATCTACTGTGAGCAAGAGAAACAGAGAAAGCCCAAGGGAGTGCACAGTGTCTATGAG 389  
QY 1292 ACCCACTGCTGCAAGGCTGCGGTGTGAGATGTTGATGGGAAAGATTTTCAAGGAGCA 1351  
DB 390 ACCCACTGCTGCAAGGCTGCGGTGTGAGATGTTGATGGGAAAGATTTTCAAGGAGCA 449  
QY 1352 AACTTAAGTCTCCCTTGTGAGAAAGAGCTCCCAATGACGATGCGGGGTGTGTGC 1411

DB 450 AACTTAAGTCTCCCTTGTGAGAAAGAGCTCCCAATGACGATGATGC-GGGTGGGCTGC 508  
QY 1412 CACCCGTGAGGAGGAGGATGAGCAGCAGCTCCGTGAGGTGACAGAGGCCAGCA 1470  
DB 509 CACCCGTGAGGAGGAGGATGAGCAGCAGCTCCGTGAGGTGACAGAGGCCAGCA 567  
RESULT 13  
US-09-864-761-7766/c  
Sequence 7766, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aomic-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 7766  
LENGTH: 505  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL162311.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1

US-09-864-761-7766

Query Match 20.3%; Score 411.6; DB 9; Length 505;  
Best Local Similarity 90.3%; Pred. No. 1.2e-109;  
Matches 486; Conservative 0; Mismatches 19; Indels 33; Gaps 3;

QY 162 AACCTATGAGAGCCCACTGATGTCACTATACCCAGGCTCAGACCACTGACCACTTATGG 221  
DB 505 AACCTATGAGAGCCCACTGATGTCACTATACCCAGGCTCAGACCACTGACCACTTATGG 446  
QY 222 GCAGACCGGCTTATGCACTTCTTATGAGACAGCTCCCACTGTTATATCTCACTCCAGTGC 281  
DB 445 G-----CTATGCACTTCTTATGAGACAGCTCCCACTGTTATATGCTCACTCCAGTGC 394  
QY 282 CCCCAGGCACTATGAGCCAGCTGTCCAGGGGTATAGGCACTGTCTTATATGATACCAAC 341  
DB 393 CCCCAGGCACTATGAGCCAGCTGTCCAGGGGTATAGGCACTGTCTTATATGATACCAAC 334  
QY 342 TGCTACAGTACCAACCAACCCAGGCTCTTATGAGCACTGATCTGATATGAGCACTGAGCC 401  
DB 333 TGCTACAGTACCAACCAACCCAGGCTCTTATGAGCACTGATCTGATATGAGCACTGAGCC 275  
QY 402 TGCTATCCAGCCTATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 461  
DB 274 TGCTATCCAGCCTATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 215  
QY 462 AAACAGCCCACTGAGACTGATCTCACTCACTTATGAGCAAGGGGTTATCAACCCAGCCAG 521  
DB 214 TAACAGCCCACTGAGACTGATCTCACTCACTTATGAGCAAGGGGTTATCAACCCAGCCAG 155  
QY 522 CTTAGATATGAGCAGAGTACTGATCTGATATCCCAAGGTTACCTGAGGAGTACCAACCCAGCA 581  
DB 154 CTTAGATATGAGCAGAGTACTGATCTGATATCCCAAGGTTACCTGAGGAGTACCAACCCAGCA 95  
QY 582 GCCAGTCACTGACCTTCATCTCACTCACTTATGAGCAAGGGGTTATCTCTACAGAGCCAGTAG 641  
DB 94 GCCAGTCACTGACCTTCATCTTCCTTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58

## RESULT 14

US-09-918-995-3868  
; Sequence 3868, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918, 995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235, 076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3868  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-3868

Query Match 19.7%; Score 399.4; DB 10; Length 401;  
Best Local Similarity 99.8%; Pred. No. 4.1e-106;  
Matches 400; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 25 GAAGGAGAGAAATGGCGTCCAGGATTAAGTATACCTATATGACCAAGTGCAGCGCAGAG 84  
DB 1 GAAGGAGAGAAATGGCGTCCAGGATTAAGTATACCTATATGACCAAGTGCAGCGCAGAG 60  
QY 85 GGCTACAGTGTCTTACACCGCCAGCCAGCTCAAGGATATGACAGACCCAGGAGTAT 144

## RESULT 15

US-09-864-761-24705  
; Sequence 24705, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hanzel, David R.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864, 761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180, 312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207, 456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632, 366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263, 6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236, 359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234, 687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608, 408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774, 203

PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
SEQ ID NO 24705  
LENGTH: 399  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC018774.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
OTHER INFORMATION: SWISSPROT HIT: Q01844, EVALUE 3.00e-28  
OTHER INFORMATION: NT HIT: g11417927, EVALUE 0.00e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BE781742.1, EVALUE 0.00e+00  
US-09-864-761-24705

Query Match 19.0%; Score 384.6; DB 9; Length 399;  
Best Local Similarity 97.7%; Pred. No. 8,7e-102;  
Matches 390; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 37 ATGGCGTCCAGGATTCAGTACCTATGCAAGCTGAGGCGAGGCGTACAGTCT 96  
DB 1 ATGGCATCCATGATTCAGTACCTATGCAAGCTGAGGCGAGGCGTACAGTCT 60  
QY 97 TACACCGCCGAGCCCACTCAAGATATGACAGACCAAGCATATGGGCAACAAGC 156  
DB 61 TACACACCCGAGCCCACTCAAGATATGACAGACCAAGCATATGGGCAACAAGC 120  
QY 157 TATGAACTATGACAGCCCACTGATGTCTATACCAAGCTCAACCACTGCAACC 216  
DB 121 TATGAACTATGACAGCCCACTGATGTCTATACCAAGCTCAACCACTGCAACC 180  
QY 217 TATGGGCAAGCCGCTTATGCACTTCTTATGACAGCCCTCCACTGGTTACTCA 276  
DB 181 TATGGGCAAGCCGCTTATGCACTTCTTATGACAGCCCTCCACTGGTTACTCA 240  
QY 277 ACTGCCCCCGAGGATTCAGGAGCTGTCCAGGGGTATGGCACTGGTCTTATGATACC 336  
DB 241 ACTGCCCCCGAGGATTCAGGAGCTGTCCAGGGGTATGGCACTGGTCTTATGATACC 300  
QY 337 ACCACTGCTACATCACCAACCAAGGCTCTATGAGCTCAGTCTGCAATATGGCACT 396  
DB 301 ACCACTGCTACATCACCAACCAAGGCTCTATGAGCTCAGTCTGCAATATGGCACT 360  
QY 397 CAGCCTGCTTATCCAGCTTATGGGCAAGCAGCAGCC 435  
DB 361 CAGCCTGCTTATCCAGCTTATGGGCAAGCAGCAGTACC 399

Search completed: February 20, 2005, 22:47:56  
Job time : 1071.67 secs





QY 61 TATAGCCAGCTGACGAGGAGGCTACAGTCTTACACCGCCAGCCCACTCAAGGA 120  
DB 65 TATAGCCAGCTGACGAGGAGGCTACAGTCTTACACCGCCAGCCCACTCAAGGA 124  
QY 121 TATGACAGACCAAGGAGGATGAGGCAACAAGCTATGAACTTATGAGACGCCACT 180  
DB 125 TATGACAGACCAAGGAGGATGAGGCAACAAGCTATGAACTTATGAGACGCCACT 184  
QY 181 GATGTGAGTATACCAAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTATGCAACT 240  
DB 185 GATGTGAGTATACCAAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTATGCAACT 244  
QY 241 TCTTATGAGACGCTCCCACTGGTTATACCTCACTCCGCCCCAGGATACAGCGAG 300  
DB 245 TCTTATGAGACGCTCCCACTGGTTATACCTCACTCCGCCCCAGGATACAGCGAG 304  
QY 301 CCTGTCCAGGGGATGAGCACTGGTCTTATGATACCACTGCTACAGTCAACCAACC 360  
DB 305 CCTGTCCAGGGGATGAGCACTGGTCTTATGATACCACTGCTACAGTCAACCAACC 364  
QY 361 CAGGCTCTTATGAGCTCAGTCTGATGAGCACTCAGCTGCTTATCCAGCTTATGGG 420  
DB 365 CAGGCTCTTATGAGCTCAGTCTGATGAGCACTCAGCTGCTTATCCAGCTTATGGG 424  
QY 421 CAGCAGCAGACGACCTGACCTTACAGACCCGAGATGGAACCAAGCCCACTGAGACT 480  
DB 425 CAGCAGCAGACGACCTGACCTTACAGACCCGAGATGGAACCAAGCCCACTGAGACT 484  
QY 481 AGTCAACCTCAATCTAGACAGGAGGATTACACCAAGCCAGCTAGGATATGAGACAGT 540  
DB 485 AGTCAACCTCAATCTAGACAGGAGGATTACACCAAGCCAGCTAGGATATGAGACAGT 544  
QY 541 AACTCAGTATATCCCAAGTACTGAGGAGCTACCCATGCAAGCACTGCACTCTCCA 600  
DB 545 AACTCAGTATATCCCAAGTACTGAGGAGCTACCCATGCAAGCACTGCACTCTCCA 604  
QY 601 TCTTACCTCTCAACGCTATCTCTCAACAGCCGCTAGTATGATTCAGACAGTATAC 660  
DB 605 TCTTACCTCTCAACGCTATCTCTCAACAGCCGCTAGTATGATTCAGACAGTATAC 664  
QY 661 TCTCAGCAGAACCTATGAGGCAACCGAGAGCTATGAGACGACGAGTATGATGATCA 720  
DB 665 TCTCAGCAGAACCTATGAGGCAACCGAGAGCTATGAGACGAGTATGATGATGATCA 724  
QY 721 CAAAGCAGTATGAGGAGGAGCTCTCCAATGATTAACCACTGATCTTACAGC 780  
DB 725 CAAAGCAGTATGAGGAGGAGCTCTCCAATGATTAACCACTGATCTTACAGC 784  
QY 781 CAAAGCTCAAGTCAATATGAGCAACAGAGCAGCTTACGAGGAGCAGAGTCACTTCCGA 840  
DB 785 CAAAGCTCAAGTCAATATGAGCAACAGAGCAGCTTACGAGGAGCAGAGTCACTTCCGA 844  
QY 841 CAGGACCAAGCTCAGTATGAGGAGGAGTCTGAGAGATTTTCCGAGCA 900  
DB 845 CAGGACCAAGCTCAGTATGAGGAGGAGTCTGAGAGATTTTCCGAGCA 904  
QY 901 GAGAGAAACCGAGAGTATGAGGAGGAGTCTGAGAGATTTTCCGAGCA 960  
DB 905 GAGAGAAACCGAGAGTATGAGGAGGAGTCTGAGAGATTTTCCGAGCA 964  
QY 961 CGTGAAGGCAATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
DB 965 CGTGAAGGCAATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1024  
QY 1021 GAGGAGAGTGGCTTCAATTAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
DB 1025 GAGGAGAGTGGCTTCAATTAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1084  
QY 1081 GGGCCACCTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
DB 1085 GGGCCACCTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1144

QY 1141 AATGACAGTGTACTCTAGATGATCTGAGAGCTTCTTTAAGCAGTGTGGGTTGTTAAG 1200  
DB 1145 AATGACAGTGTACTCTAGATGATCTGAGAGCTTCTTTAAGCAGTGTGGGTTGTTAAG 1204  
QY 1201 ATGAACAGAGAACTGGGCAACCTATGATCCATCTTACCTTGAACAGAGAAACAGAGAAAG 1260  
DB 1205 ATGAACAGAGAACTGGGCAACCTATGATCCATCTTACCTTGAACAGAGAAACAGAGAAAG 1264  
QY 1261 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
DB 1265 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1324  
QY 1321 TGGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
DB 1325 TGGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1384  
QY 1381 CCTCCATGAAACAGTATGCGGGGTGTCTGCAACCCGAGAGGAGGAGGAGGAGGAGGAGGAG 1440  
DB 1385 CCTCCATGAAACAGTATGCGGGGTGTCTGCAACCCGAGAGGAGGAGGAGGAGGAGGAGGAG 1444  
QY 1441 CCACTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
DB 1445 CCACTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1504  
QY 1501 GGCCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
DB 1505 GGCCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1564  
QY 1561 CCTCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620  
DB 1565 CCTCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1624  
QY 1621 TGTGAGAAACAGAACTTGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
DB 1625 TGTGAGAAACAGAACTTGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1684  
QY 1681 GAAAGCTTCTCCGCAACCTTTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
DB 1685 GAAAGCTTCTCCGCAACCTTTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1744  
QY 1741 GGTGCAATCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
DB 1745 GGTGCAATCCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1804  
QY 1801 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860  
DB 1805 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1864  
QY 1861 GGCCTTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920  
DB 1865 GGCCTTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1924  
QY 1921 ATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
DB 1925 ATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1984  
QY 1981 CAGGAGCGGAGAGATGCGCTTACTAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2025  
DB 1985 CAGGAGCGGAGAGATGCGCTTACTAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2029

RESULT 2  
CR608375  
LOCUS CR608375 2314 bp mRNA linear HTC 21-JUN-2004  
DEFINITION full-length cDNA clone CS0D1037J24 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR608375  
VERSION CR608375.1 GI:50489182  
KEYWORDS HTC; cDNA; cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.





QY	1201	ATGGAACAAGAGACTGGGGCAACCCATGATCCACATCTACCTGGACAACAAGAAACAGAAAG	1266
Db	1207	ATGAACAGAGAACTGGGGCAACCCATGATCCACATCTACCTGGACAACAAGAAACAGAAAG	1266
QY	1261	CCCAAGGCGATGCCACAGTGTCTATGAAAGACCAACCACTGCACAAAGCTGCCTGGAA	1320
Db	1267	CCCAAGGCGATGCCACAGTGTCTATGAAAGACCAACCACTGCACAAAGCTGCCTGGAA	1320
QY	1321	TGGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCTTGTCTGGAAAG	1380
Db	1327	TGGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCTTGTCTGGAAAG	1380
QY	1381	CCTCAATGAACAGATATGGGGGTGTCTGCAACCCCGTGAAGGGGAGAGCATGCAACA	1440
Db	1387	CCTCAATGAACAGATATGGGGGTGTCTGCAACCCCGTGAAGGGGAGAGCATGCAACA	1440
QY	1441	CCACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTCTGGGGAGACCCATGGGTGCAATGGGA	1500
Db	1447	CCACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTCTGGGGAGACCCATGGGTGCAATGGGA	1500
QY	1501	GGCCCTGAGAGGATATGAGAGAGCTTCCCTCCAAAGACCCCGGGGTTCCGAGGGAAAC	1560
Db	1507	GGCCCTGAGAGGATATGAGAGAGCTTCCCTCCAAAGACCCCGGGGTTCCGAGGGAAAC	1560
QY	1561	CCCTTGGAGAGGAAACGTCCAGACACGAGCTGGAGACTGGCAAGTGTCCCAATCCGGGT	1620
Db	1567	CCCTTGGAGAGGAAACGTCCAGACACGAGCTGGAGACTGGCAAGTGTCCCAATCCGGGT	1620
QY	1621	TGTGAAACCAAGAACTTGCTTGAGAAACAAGTGCACAGATGTAAAGGCCCAAGGCT	1680
Db	1627	TGTGAAACCAAGAACTTGCTTGAGAAACAAGTGCACAGATGTAAAGGCCCAAGGCT	1680
QY	1681	GAAAGCTTCTCCCGCCACCTTTCCGGCCCCGGGTGGTATCTGTGCAGAGGTGGCCCT	1740
Db	1687	GAAAGCTTCTCCCGCCACCTTTCCGGCCCCGGGTGGTATCTGTGCAGAGGTGGCCCT	1740
QY	1741	GTTGGCATCGGGGAGAAAGAGTGGCTCATGATCTGTGGTGGCCCCGGTGAATGTTC	1800
Db	1747	GTTGGCATCGGGGAGAAAGAGTGGCTCATGATCTGTGGTGGCCCCGGTGAATGTTC	1800
QY	1801	AGAGGTGCCGTGTGTGAGACAGAGTGGCTTCCGTGTGGCCCGGGGAGTGAACCGAGT	1860
Db	1807	AGAGGTGCCGTGTGTGAGACAGAGTGGCTTCCGTGTGGCCCGGGGAGTGAACCGAGT	1860
QY	1861	GGCTTTGGTGGAGAAAGACAGAGTGGCCCTTGGGGGGCCCCCTTGAACCTTTGATGGAACAG	1920
Db	1867	GGCTTTGGTGGAGAAAGACAGAGTGGCCCTTGGGGGGCCCCCTTGAACCTTTGATGGAACAG	1920
QY	1921	ATGGAGAGAAAGAGAGAGAGCTGTGAAGACCTGTGAATAATGATTAAGAGGAGACCGT	1980
Db	1927	ATGGAGAGAAAGAGAGAGAGCTGTGAAGACCTGTGAATAATGATTAAGAGGAGACCGT	1980
QY	1981	CAGAGCGCAGAGATCGGCTCTACTAGTCAAGAGACCCCGCAGA	2025
Db	1987	CAGAGCGCAGAGATCGGCTCTACTAGTCAAGAGACCCCGCAGA	2025

Query Match	Best Local Similarity	Matches 1977; Conservative	Score 1977; DB 3; Length 2103;
49	GATTACAGTACTCTTATGCGCAGGCTGCGACGCGACGAGCGGGCTACAGTCTTACACCGGCCGAC	100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;	108
1	GATTACAGTACTCTTATGCGCAGGCTGCGACGCGACGAGCGGGCTACAGTCTTACACCGGCCGAC		60
109	CCCATCTCAGGATATGACACGACCCAGGCGATATGCGCAACAAAGCTATGGAACCTAT		168
61	CCCATCTCAGGATATGACACGACCCAGGCGATATGCGCAACAAAGCTATGGAACCTAT		120
169	GGAACGCCCATCTGATGTCAGCTTATACCCAGGCTCAGACCACTGCAACCTATGCGAGACC		228
121	GGAACGCCCATCTGATGTCAGCTTATACCCAGGCTCAGACCACTGCAACCTATGCGAGACC		180
229	GCCATATGGAATTTCTTATGACACGCTTCCCACTGGTATATCTATCTCCAACTGCCCCCAG		288
181	GCCATATGGAATTTCTTATGACACGCTTCCCACTGGTATATCTATCTCCAACTGCCCCCAG		240
289	GCATACAGCCAGCGCTGTCAGAGGGGATGAGCACTGAGTATGATACCAACCACTGCTTAC		348
241	GCATACAGCCAGCGCTGTCAGAGGGGATGAGCACTGAGTATGATACCAACCACTGCTTAC		300
349	GTCACACACACCGAGGCTCTTATGACGCTTGTGATATGAGCACTGAGCTCAGCTGCTTAT		408
301	GTCACACACACCGAGGCTCTTATGACGCTTGTGATATGAGCACTGAGCTCAGCTGCTTAT		360
409	CCAGCTTATGCGGCGACGACGCGACGCGACCTTACACGACCGGAGATGGAACAAG		468
361	CCAGCTTATGCGGCGACGACGCGACGCGACCTTACACGACCGGAGATGGAACAAG		420
469	CCCATCTGAGATTAATCAACCTCATTTAGACACAGGGGGTTTACAACGACCGAGCTTACGA		528
421	CCCATCTGAGATTAATCAACCTCATTTAGACACAGGGGGTTTACAACGACCGAGCTTACGA		480
529	TATGACAGAGTAACTACAGTTATCCCGAGGTAAGTGGAGCTTACCCCATGACGACGTC		588
481	TATGACAGAGTAACTACAGTTATCCCGAGGTAAGTGGAGCTTACCCCATGACGACGTC		540
589	ACTGCACTCTCATCTTACCTCTTACCAAGCTATTTCTTACACGACCGACTAGTTATGAT		648
541	ACTGCACTCTCATCTTACCTCTTACCAAGCTATTTCTTACACGACCGACTAGTTATGAT		600
649	CAGAGCAGTAACTCTTACAGCAAGAAACCTATGAGGGAACGAGCACTTATGACACAGT		708
601	CAGAGCAGTAACTCTTACAGCAAGAAACCTATGAGGGAACGAGCACTTATGACACAGT		660

QY 709 AGCTATGTCACCAAGAGCTATGAGGAGCAGCTCCCACTAGTATACCAACCCCAACT 768  
 DB 661 AGCTATGTCACCAAGAGCTATGAGGAGCAGCTCCCACTAGTATACCAACCCCAACT 720  
 QY 769 GATTCCTACAGCCAGCTCCCACTAGTATACCAAGAGCAGCTACCGGAGCAG 828  
 DB 721 GATTCCTACAGCCAGCTCCCACTAGTATACCAAGAGCAGCTACCGGAGCAG 780  
 QY 829 AGTTCATTCGACAGAGCAGCCCACTAGTATACCGGAGCAGCTACCGGAGCAG 888  
 DB 781 AGTTCATTCGACAGAGCAGCCCACTAGTATACCGGAGCAGCTACCGGAGCAG 840  
 QY 889 TTTCCGAGCAGAGAGAGCAGCCCACTAGTATACCGGAGCAGCTACCGGAGCAG 948  
 DB 841 TTTCCGAGCAGAGAGAGCAGCCCACTAGTATACCGGAGCAGCTACCGGAGCAG 900  
 QY 949 GGGGGATTTGATCGTGGAGGATGAGAGAGGTTGGGGAGAGAGCAGCGGTGGAATG 1008  
 DB 901 GGGGGATTTGATCGTGGAGGATGAGAGAGGTTGGGGAGAGAGCAGCGGTGGAATG 960  
 QY 1009 GGCAGCGCTGAGAGAGGAGGTTGCTTCAATTAACCTGTTGAGACCCATGATGAAGACCA 1068  
 DB 961 GGCAGCGCTGAGAGAGGAGGTTGCTTCAATTAACCTGTTGAGACCCATGATGAAGACCA 1020  
 QY 1069 GATCTTGATCTAGGAGCCCACTGATAGATCAGATGAAGACTCTGACCAACAGTGCATTTAT 1128  
 DB 1021 GATCTTGATCTAGGAGCCCACTGATAGATCAGATGAAGACTCTGACCAACAGTGCATTTAT 1080  
 QY 1129 GTACAGAGATTAATAGACAGTGTGACTCTAGATGATCTGAGCAGACTCTTTAAAGCAGTGT 1188  
 DB 1081 GTACAGAGATTAATAGACAGTGTGACTCTAGATGATCTGAGCAGACTCTTTAAAGCAGTGT 1140  
 QY 1189 GGGGTTTGAATGAACAG 1248  
 DB 1141 GGGGTTTGAATGAACAG 1200  
 QY 1249 GAAACAG 1308  
 DB 1201 GAAACAG 1260  
 QY 1309 GCTGCGCTGAGATGTTTGAAG 1368  
 DB 1261 GCTGCGCTGAGATGTTTGAAG 1320  
 QY 1369 GCTGCGAGAGAGAGCTCCCAATGAACAGTATGCGGGGTGCTTCCCAAGAGAGAGAGAGAGAG 1428  
 DB 1321 GCTGCGAGAGAGAGCTCCCAATGAACAGTATGCGGGGTGCTTCCCAAGAGAGAGAGAGAGAG 1380  
 QY 1429 GGCATGACACCAACACTCGGTGAG 1488  
 DB 1381 GGCATGACACCAACACTCGGTGAG 1440  
 QY 1489 GGTGCGATGAGAGAGAGCTGAG 1548  
 DB 1441 GGTGCGATGAGAGAGAGCTGAG 1500  
 QY 1549 TCCCGAGAGAGAGAGCTCTGAG 1608  
 DB 1501 TCCCGAGAGAGAGAGCTCTGAG 1560  
 QY 1609 CCCAATCGGGTGTGAG 1668  
 DB 1561 CCCAATCGGGTGTGAG 1620  
 QY 1669 GCCCAAGAGAGAGAGCTCTTCCGAG 1728  
 DB 1621 GCCCAAGAGAGAGAGCTCTTCCGAG 1680  
 QY 1729 AGAGGTGGCTGTGAG 1788  
 DB 1681 AGAGGTGGCTGTGAG 1740  
 QY 1789 GGTGGAATGTTCAAG 1848

DB 1741 GGTGGAATGTTCAAG 1800  
 QY 1849 ATGAGACCAAGAGAGAGCTTTGATGAG 1908  
 DB 1801 ATGAGACCAAGAGAGAGCTTTGATGAG 1860  
 QY 1909 TTGATGAACAGATGAG 1968  
 DB 1861 TTGATGAACAGATGAG 1920  
 QY 1969 GGCAG 2025  
 DB 1921 GGCAG 1977  
 RESULT 5  
 CR602561  
 LOCUS  
 DEFINITION  
 full-length cDNA clone CS01068YN23 of Placenta Cot 25-normalized  
 of Homo sapiens (human).  
 ACCESSION  
 CR602561.1 GI:50483368  
 VERSION  
 CR602561.1  
 KEYWORDS  
 HTC; CNSLT\_cDNA.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 AUTHORS  
 Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE  
 Full-length cDNA libraries and normalization  
 JOURNAL  
 Unpublished  
 REMARK  
 Contact : Feng Liang Email : fliang@life.rockefeller.edu  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue  
 Redwood City, CA 94063  
 2 (bases 1 to 2070)  
 REFERENCE  
 AUTHORS  
 Genoscope.  
 TITLE  
 Direct Submision  
 JOURNAL  
 Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT  
 - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES  
 source  
 1..2070  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS01068YN23"  
 /tissue\_type="Placenta. Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 96.8%; Score 1961; DB 3; Length 2070;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1974; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 49 GATTACAGTACCTATGAG 108  
 DB 1 GATTACAGTACCTATGAG 60  
 QY 109 CCCACTCAAGATATGAG 168  
 DB 61 CCCACTCAAGATATGAG 120  
 QY 169 GAG 228  
 DB 121 GAG 180  
 QY 229 GCTATGCAACTTCTATGAG 288

Db 181 GCTTATGCACTTCTTATGAGCAGCTTCCCTGCTTATCTACTCTCACTTGGCCCCAG 240  
 Qy 289 GCATACAGCCAGCCTGTCCAGGGGTATGCACTGCTGCTTATGATACCACTGCTACA 348  
 Db 241 GCATACAGCCAGCCTGTCCAGGGGTATGCACTGCTGCTTATGATACCACTGCTACA 300  
 Qy 349 GTCAACCAACCCAGGCTCTTATGCACTGCTGCTGCTTATGATACCACTGCTACA 408  
 Db 301 GTACCAACCAACCCAGGCTCTTATGCACTGCTGCTGCTTATGATACCACTGCTACA 360  
 Qy 409 CCAGCTTATGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 468  
 Db 361 CCAGCTTATGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420  
 Qy 469 CCCACTGAGACTAGTCAACTCTATGCACTGAGGAGGTTACACAGCCAGCCTTATGA 528  
 Db 421 CCCACTGAGACTAGTCAACTCTATGCACTGAGGAGGTTACACAGCCAGCCTTATGA 480  
 Qy 529 TATGACAGAGTAACTACAGTTATCCCAAGGTACCTGAGGAGTACCCCACTGAGCCAGT 588  
 Db 481 TATGACAGAGTAACTACAGTTATCCCAAGGTACCTGAGGAGTACCCCACTGAGCCAGT 540  
 Qy 589 ACTGACCTTCATCTTACCTCTACAGGCTATTCCTCTACAGAGCCAGCTATGAT 648  
 Db 541 ACTGACCTTCATCTTACCTCTACAGGCTATTCCTCTACAGAGCCAGCTATGAT 600  
 Qy 649 CAGAGCAGTTACTCTCAGAGAAACCTATGAGGCAACGAGCAGCTATGACAGCAGAGT 708  
 Db 601 CAGAGCAGTTACTCTCAGAGAAACCTATGAGGCAACGAGCAGCTATGACAGCAGAGT 660  
 Qy 709 AGCTATGCTCAACAAAGCAGCTATGAGGAGCAGCTTCCCTAGTTATCCCACTGAGT 768  
 Db 661 AGCTATGCTCAACAAAGCAGCTATGAGGAGCAGCTTCCCTAGTTATCCCACTGAGT 720  
 Qy 769 GGATCTCAACAGCAGCAGCTCAATATAGCAACAGAGCAGCAGCTACGGGAGCAG 828  
 Db 721 GGATCTCAACAGCAGCAGCTCAATATAGCAACAGAGCAGCAGCTACGGGAGCAG 780  
 Qy 829 AGTTCACTCCGAGAGCAACCCAGTATGAGTGGGTATGAGGAGAGTCTGAGGA 888  
 Db 781 AGTTCACTCCGAGAGCAACCCAGTATGAGTGGGTATGAGGAGAGTCTGAGGA 840  
 Qy 889 TTTTCCGAGCAGAGAGAAACCGAGATGAGTGGCTCTGATACCGGGGAGGGGAGA 948  
 Db 841 TTTTCCGAGCAGAGAGAAACCGAGATGAGTGGCTCTGATACCGGGGAGGGGAGA 900  
 Qy 949 GGGGATTTGATGATGAGGAGCATGAGCAGAGTGGGAGGGAGAGAGAGCGGGTGAATG 1008  
 Db 901 GGGGATTTGATGATGAGGAGCATGAGCAGAGTGGGAGGGAGAGAGAGCGGGTGAATG 960  
 Qy 1009 GGGAGGCTGAGAGAGCGAGGTGCTTCAATATAGCCTGCTGAGCAGCAGTATGAGAGCA 1068  
 Db 961 G---GGGCTGAGAGAGCGAGGTGCTTCAATATAGCCTGCTGAGCAGCAGTATGAGAGCA 1017  
 Qy 1069 GATCTTGAATTAGCCCACTGTATGATCCAGTATGAGAGCTCTGACCAAGTGAATTTAT 1128  
 Db 1018 GATCTTGAATTAGCCCACTGTATGATCCAGTATGAGAGCTCTGACCAAGTGAATTTAT 1077  
 Qy 1129 GTACAGAGATTAAATGCACTGATGATCTGATGATCTGACAGCTTCTTATAGCAGTGT 1188  
 Db 1078 GTACAGAGATTAAATGCACTGATGATCTGATGATCTGACAGCTTCTTATAGCAGTGT 1137  
 Qy 1189 GGGGTTTAAATGATGAGCAAGAGAACTGGCAACCCATGATCACTTACTGAGCAAG 1248  
 Db 1138 GGGGTTTAAATGATGAGCAAGAGAACTGGCAACCCATGATCACTTACTGAGCAAG 1197  
 Qy 1249 GAAACAGAGAAACCCAGAGCGATGCACTGTCTTATGAGAGCCCACTGAGCAAG 1308  
 Db 1198 GAAACAGAGAAACCCAGAGCGATGCACTGTCTTATGAGAGCCCACTGAGCAAG 1257  
 Qy 1309 GCTGCGTGAATGTTGATGAGAAAGATTTTCAAGGGAGCAACTTAAAGTCTCCCTT 1368

Db 1258 GCTGCCGTGGAATGCTTTGATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTT 1317  
 Qy 1369 GCTCGAAGAAAGCTTCCAAATGAACAGTATGCGGGGTGTGTCACCCGCTGAGGGAGAGA 1428  
 Db 1318 GCTCGAAGAAAGCTTCCAAATGAACAGTATGCGGGGTGTGTCACCCGCTGAGGGAGAGA 1377  
 Qy 1429 GGCATGCCAACCACTTCCCTGAGAGTCCAGAGAGCCCAAGAGAGTCTTGGGGAGACCATG 1488  
 Db 1378 GGCATGCCAACCACTTCCCTGAGAGTCCAGAGAGCCCAAGAGAGTCTTGGGGAGACCATG 1437  
 Qy 1489 GGTGCAATGGGAGGCGCTGAGAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCGGGGGT 1548  
 Db 1438 GGTGCAATGGGAGGCGCTGAGAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCGGGGGT 1497  
 Qy 1549 TCCGAGAGGAAACCTCTTGTGAGAGAGAAAGTCCAGACCCAGCTGAGACTGAGT 1608  
 Db 1498 TCCGAGAGGAAACCTCTTGTGAGAGAGAAAGTCCAGACCCAGCTGAGACTGAGT 1557  
 Qy 1609 CCCAATCCGGGTGTGAGAAACCAAGAACTTGGCTGTGAGAAACAGATGCAACAGTGTAG 1668  
 Db 1558 CCCAATCCGGGTGTGAGAAACCAAGAACTTGGCTGTGAGAAACAGATGCAACAGTGTAG 1617  
 Qy 1669 GCCCAAGAGCTGAGAGGCTTCTCCGCGCAACCTTCCGCGCCCGGGTGTGATCGTGGC 1728  
 Db 1618 GCCCAAGAGCTGAGAGGCTTCTCCGCGCAACCTTCCGCGCCCGGGTGTGATCGTGGC 1677  
 Qy 1729 AGAGTGGCCCTGTGTGAGATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGTCC 1788  
 Db 1678 AGAGTGGCCCTGTGTGAGATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGTGCC 1737  
 Qy 1789 GGTGGAATGTTCAAGAGTGGCCCTGTGTGAGAAACAGATGAGTGGCTTCCGTGTGTGCC 1848  
 Db 1738 GGTGGAATGTTCAAGAGTGGCCCTGTGTGAGAAACAGATGAGTGGCTTCCGTGTGTGCC 1797  
 Qy 1849 ATGACCAAGAGTGGCTTGTGTGAGAGAAAGCAGAGTGGCCCTGTGGGGGCCCCCTGAGCT 1908  
 Db 1798 ATGACCAAGAGTGGCTTGTGTGAGAGAAAGCAGAGTGGCCCTGTGGGGGCCCCCTGAGCT 1857  
 Qy 1909 TTGATGGAACAGATGAT 1968  
 Db 1858 TTGATGGAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1917  
 Qy 1969 GGGGAGCAGCTGTGAGAGAGCAGAGATCGGCCCTTACTAGTACAGAGAGAGAGAGAG 2025  
 Db 1918 GGGGAGCAGCTGTGAGAGAGCAGAGATCGGCCCTTACTAGTACAGAGAGAGAGAGAG 1974

RESULT 6  
 CR594868  
 LOCUS 2093 bp mRNA linear HTC 21-JUL-2004  
 DEFINITION full-length cDNA clone CS0D1064YE15 of Placenta Cot 25-normalized  
 of Homo sapiens (human).  
 ACCESSION CR594868  
 VERSION CR594868.1 GI:50475675  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2093)  
 AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polyes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue  
 2 (bases 1 to 2093)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submission  
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr  
 JOURNAL - Web : www.genoscope.cns.fr)  
 COMMENT 1st strand cDNA was primed with a Nott1-oligo(dT) primer. Five prime



end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES  
source  
1. 2093  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS01064YE15"  
/tissue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 96.8%; Score 1961; DB 3; Length 2093;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1974; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

49 GATTACAGTACCTATGCGAAGCTGACAGGCGAGGCGCTACAGTGTCTTACACCGCCAG 108  
1 GATTACAGTACCTATGCGAAGCTGACAGGCGAGGCGCTACAGTGTCTTACACCGCCAG 60  
109 CCCACTCAAGATATGACAGACACCCAGGCAATATGGGCAACAAAGCTATGGAACCTAT 168  
61 CCCACTCAAGATATGACAGACACCCAGGCAATATGGGCAACAAAGCTATGGAACCTAT 120  
169 GGAACAGCCCATATGTCAGCTATACCCAGGCTCAAGACCTGCAACCTATGGGAGACC 228  
121 GGAACAGCCCATATGTCAGCTATACCCAGGCTCAAGACCTGCAACCTATGGGAGACC 180  
229 GCGTATGCAACTCTTATGACAGACCTCCCACTGTTATCTCACTGCCCCCAG 288  
181 GCGTATGCAACTCTTATGACAGACCTCCCACTGTTATCTCACTGCCCCCAG 240  
289 GCATACAGCCAGCCTGTCAGGGGATATGACCTGTGTCTTATGATACCACTGCTTAC 348  
241 GCATACAGCCAGCCTGTCAGGGGATATGACCTGTGTCTTATGATACCACTGCTTAC 300  
349 GTACACACCCAGCCTGTCAGGGGATATGACCTGTGTCTTATGATACCACTGCTTAC 408  
301 GTACACACCCAGCCTGTCAGGGGATATGACCTGTGTCTTATGATACCACTGCTTAC 360  
409 CCAGCCTATGGGAGAGCAGCAGACCACTGCACTTACAGACCGAGATGAGAAACAG 468  
361 CCAGCCTATGGGAGAGCAGCAGACCACTGCACTTACAGACCGAGATGAGAAACAG 420  
469 CCCACTGAGACTATGCACTTATGACAGGGGGTTTAAACCAAGCCAGCCTTATGGA 528  
421 CCCACTGAGACTATGCACTTATGACAGGGGGTTTAAACCAAGCCAGCCTTATGGA 480  
529 TATGACAGAGTAACTTACATTTATCCCAAGTACCTGGAGCTTATCCCATGAGCCATG 588  
481 TATGACAGAGTAACTTACATTTATCCCAAGTACCTGGAGCTTATCCCATGAGCCATG 540  
589 ACTGACCTCATCTCTACCTCTCTACAGAGCTTATCTCTACAGCCGCACTAAGTTATG 648  
541 ACTGACCTCATCTCTACCTCTCTACAGAGCTTATCTCTACAGCCGCACTAAGTTATG 600  
649 CAGAGCAGTTACTTCTGAGCAACCACTATGGGCAACCGAGCAGCTATGAGCAGAGAGT 708  
601 CAGAGCAGTTACTTCTGAGCAACCACTATGGGCAACCGAGCAGCTATGAGCAGAGAGT 660  
709 AGCTATGCTTAAACAAAGCAGCTATGGGCAACCGAGCAGCTTATCCCACTTAACT 768  
661 AGCTATGCTTAAACAAAGCAGCTATGGGCAACCGAGCAGCTTATCCCACTTAACT 720  
769 GATTCCTACAGCAAGCTTCAAGTCAATATAGCAACAGAGCAGCAGCTGAGGAGCAG 828  
721 GATTCCTACAGCAAGCTTCAAGTCAATATAGCAACAGAGCAGCAGCTGAGGAGCAG 780  
829 AGTTCATTCGAGCAGACACCCAGATGAGATGGGTTTATATGGCAGAGAGTCTGAGGA 888  
781 AGTTCATTCGAGCAGACACCCAGATGAGATGGGTTTATATGGCAGAGAGTCTGAGGA 840

889 TTTTCGAGCCAGAGAGAACCGAGCATGATGAGCCCTGATTAACCGGGCAGGGAGAGA 948  
841 TTTTCGAGCCAGAGAGAACCGAGCATGATGAGCCCTGATTAACCGGGCAGGGAGAGA 900  
949 GGGGAGATTTGATCGTGAAGCATGACAGAGGTGGCGGGAGAGAGACCGGTGGAATG 1008  
901 GGGGAGATTTGATCGTGAAGCATGACAGAGGTGGCGGGAGAGAGACCGGTGGAATG 960  
1009 GGCAGCGCTGAGAGAGCAGAGTGGCTTCAATAGCCTGTGTGAGACCTTGAATAGAGCA 1068  
961 G--GCGCTGAGAGAGCAGAGTGGCTTCAATAGCCTGTGTGAGACCTTGAATAGAGCA 1017  
1069 GATCTTGAATCTAGGCGCCACCTGTAGATCCAGATGAAGACTCTGACAAAGTGAATTTAT 1128  
1018 GATCTTGAATCTAGGCGCCACCTGTAGATCCAGATGAAGACTCTGACAAAGTGAATTTAT 1077  
1129 GTACAAAGATTAATGACAGTGTGACTTGAATGATCTGGCAGACTTCTTTAAGCAGTGT 1188  
1078 GTACAAAGATTAATGACAGTGTGACTTGAATGATCTGGCAGACTTCTTTAAGCAGTGT 1137  
1189 GGGGTGTTAAGATGAACAAGAACTGGGCAACCCATGATCACTTATCCTGAGCAAG 1248  
1138 GGGGTGTTAAGATGAACAAGAACTGGGCAACCCATGATCACTTATCCTGAGCAAG 1197  
1249 GAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAAGACCACTGACCAAG 1308  
1198 GAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAAGACCACTGACCAAG 1257  
1309 GCTGCGGTGAAATGTTTGAATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCT 1368  
1258 GCTGCGGTGAAATGTTTGAATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCT 1317  
1369 GCTGCGAAAGACCTTCCATTAAGACATATGCGGGGTGTCTGCAACCCGTGAGGGCAGA 1428  
1318 GCTGCGAAAGACCTTCCATTAAGACATATGCGGGGTGTCTGCAACCCGTGAGGGCAGA 1377  
1429 GGCATGCAACCACTCCGTGAGAGGTCCAGAGGCGCCAGAGGTCTTGGGGAGACCATG 1488  
1378 GGCATGCAACCACTCCGTGAGAGGTCCAGAGGCGCCAGAGGTCTTGGGGAGACCATG 1437  
1489 GGTGCGATGGAGGCGGTGAGAGATGAGAGGCTTCCCTTCAAGAGAACCCCGGGGT 1548  
1438 GGTGCGATGGAGGCGGTGAGAGATGAGAGGCTTCCCTTCAAGAGAACCCCGGGGT 1497  
1549 TCCGAGGGAACCCCTCTGAGAGAGAAACCTCCAGCAGCAGCTGAGACCTGGCAGTGT 1608  
1498 TCCGAGGGAACCCCTCTGAGAGAGAAACCTCCAGCAGCAGCTGAGACCTGGCAGTGT 1557  
1609 CCCAATCCGGGTTGTGAAACCAAGAACTTCCGCTGAGAAACAGATGCAACCAAGTAAAG 1668  
1558 CCCAATCCGGGTTGTGAAACCAAGAACTTCCGCTGAGAAACAGATGCAACCAAGTAAAG 1617  
1669 GCCCAAGACCTGAAGGCTTCTCCGCGCACTTTCGCGCCCGGGGTGTATCGTGC 1728  
1618 GCCCAAGACCTGAAGGCTTCTCCGCGCACTTTCGCGCCCGGGGTGTATCGTGC 1677  
1729 AAGAGTGGCCCTGTGTGAGATGCGGGGAGAGAAAGTGGCTTATGATCGTGTGTGCC 1788  
1678 AAGAGTGGCCCTGTGTGAGATGCGGGGAGAGAAAGTGGCTTATGATCGTGTGTGCC 1737  
1789 GGTGGAATGTTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGTGCGGGG 1848  
1738 GGTGGAATGTTTCAAGAGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGTGCGGGG 1797  
1849 ATGACCGAGGTGGCTTGTGTGAGAGAAAGAGTGGCTTGTGGGGGCGCCCTGTGACCT 1908  
1798 ATGACCGAGGTGGCTTGTGTGAGAGAAAGAGTGGCTTGTGGGGGCGCCCTGTGACCT 1857  
1909 TTGATGAAACAGATGGAGAGAAAGAGAGACGTGTGAGAGACCTTGAATAAATGATATAA 1968  
1858 TTGATGAAACAGATGGAGAGAAAGAGAGACGTGTGAGAGACCTTGAATAAATGATATAA 1917



Db 1438 GGTGGATGGAGGCGCTGGAGAGATAGAGGCTTCCCTCCAGAGAGACCCCGGGT 1497  
 1549 TCCCGAGGAAACCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGAGTGGAGTGT 1608  
 1498 TCCCGAGGAAACCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGAGTGGAGTGT 1557  
 1609 CCCAATCCGGGTGTTGAGAAACGAACTTCGCTGAGAAACAGAGTGAACCGAGTAA 1668  
 1558 CCCAATCCGGGTGTTGAGAAACGAACTTCGCTGAGAAACAGAGTGAACCGAGTAA 1617  
 1669 GCGCCAAAGCTGAGAGGCTTCTCCCGCCACCTTTCCGCCCCGGGGTGGTGTGCT 1728  
 1618 GCGCCAAAGCTGAGAGGCTTCTCCCGCCACCTTTCCGCCCCGGGGTGGTGTGCT 1677  
 1729 AGAGTGGCCCTGTTGAGAGTGGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1788  
 1678 AGAGTGGCCCTGTTGAGAGTGGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1737  
 1789 GGTGAGATGTTGAGAGTGGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1848  
 1738 GGTGAGATGTTGAGAGTGGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1797  
 1849 ATGACCGAGTGGCTTCTGAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1908  
 1798 ATGACCGAGTGGCTTCTGAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1857  
 1909 TTGATGAGACGATGAGAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1968  
 1858 TTGATGAGACGATGAGAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1917  
 1969 GGGAGACCGTGAAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 2025  
 1918 GGGAGACCGTGAAGAGAGAGAGAGAGTGGCTTCTGAGTGGTGTGCT 1974

RESULT 8  
 CR593724  
 LOCUS full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized  
 DEFINITION of Homo sapiens (human).  
 CR593724  
 CR593724.1 GI:50474531  
 VERSION HTC; cDNA.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 2073)  
 L1.W.B., Gruber, C., Jessee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
 2 (bases 1 to 2073)  
 Genoscope.  
 Direct Submission  
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 101 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 COMMENT  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 REMARK  
 FEATURES  
 source  
 1..2073  
 /organism="Homo sapiens"  
 /mol\_type="cDNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1011Y005"  
 /tissue\_type="Placenta Cot 25-normalized"

ORIGIN /plasmid="pCMVSPORT\_6"  
 Query Match 95.8%; Score 1940; DB 3; Length 2073;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 86 GGTACAGTCTTACACCGCCAGCCCACTGAGATATACAGACCAACCCAGGATATG 145  
 1 GGTACAGTCTTACACCGCCAGCCCACTGAGATATACAGACCAACCCAGGATATG 60  
 146 GGCACAAAGCTATGGAACCTATGAGACAGCCCACTGATGTCACTATACCGAGCT 205  
 61 GGCACAAAGCTATGGAACCTATGAGACAGCCCACTGATGTCACTATACCGAGCT 120  
 206 CCACTGCAACCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265  
 121 CCACTGCAACCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 266 ATACTACTCAACTGCCCCCGAGGATACAGAGAGAGAGAGAGAGAGAGAGAGAG 325  
 181 ATACTACTCAACTGCCCCCGAGGATACAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 326 CTTATGATACCACTGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 385  
 241 CTTATGATACCACTGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 300  
 386 CATATGAGCACTAGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 445  
 301 CATATGAGCACTAGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 360  
 446 CAAGACCGAGAGTGAACCAAGCCCACTGAGACTAGTCAACTCACTATGAGACAG 505  
 361 CAAGACCGAGAGTGAACCAAGCCCACTGAGACTAGTCAACTCACTATGAGACAG 420  
 506 GTTACAAAGCTGAGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 565  
 421 GTTACAAAGCTGAGCTGCTTATCCAGCTTATGAGAGAGAGAGAGAGAGAGAG 480  
 566 GAGAGTACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625  
 481 GAGAGTACCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 626 CTACACAGCTGAGTGTATGATCAGAGAGAGTACTCTGAGAGAGAGAGAGAGAG 685  
 541 CTACACAGCTGAGTGTATGATCAGAGAGAGTACTCTGAGAGAGAGAGAGAGAG 600  
 686 CGAGAGCTATGAG 745  
 601 CGAGAGCTATGAG 660  
 746 CCACTGTTATCCACCCCAACTGATCTTACAGAGAGAGAGAGAGAGAGAGAGAG 805  
 661 CCACTGTTATCCACCCCAACTGATCTTACAGAGAGAGAGAGAGAGAGAGAGAG 720  
 806 AGAGAGAGCTACAG 865  
 721 AGAGAGAGCTACAG 780  
 866 TTTATGAG 925  
 781 TTTATGAG 840  
 926 CTGATTAACCGGAG 985  
 841 CTGATTAACCGGAG 900  
 986 GGGAG 1045  
 901 GGGAG 960  
 1046 GTGAGACCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105

Db 961 GTGAGCCATGATGAGACACAGATCTTATCTAGGCCACCTGTAGATCCAGATGAG 1020  
 QY 1106 ACTCTGACAACTGATGATTTATGTAACAAGATTTAAATGCACTGTGACTGTAGATC 1165  
 Db 1021 ACTCTACACAGTGTGAAATTTATGTAACAGATTTAAATGCACTGTGACTGTAGATC 1080  
 QY 1166 TGGCAGACTCTTTAAAGCAGTGTGGGGTGTTAAGATGAAACAAGAACTGGGCAACCC 1225  
 Db 1081 TGGCAGACTCTTTAAAGCAGTGTGGGGTGTTAAGATGAAACAAGAACTGGGCAACCC 1140  
 QY 1226 TGATCCACTTACTCTGAGACAGAAACAGAAAGCCCAAGGCGATGCCAGATGTCT 1285  
 Db 1141 TGATCCACTTACTCTGAGACAGAAACAGAAAGCCCAAGGCGATGCCAGATGTCT 1200  
 QY 1286 ATGAAACCCACCCACTGCGCAAGGCTGCGCTGTGAATGTGTTGATGGGAAAGATTTTCA 1345  
 Db 1201 ATGAAACCCACCCACTGCGCAAGGCTGCGCTGTGAATGTGTTGATGGGAAAGATTTTCA 1260  
 QY 1346 GGAGCAAACTTAAAGTCTCCCTTGTCTCGGAAAGAGCTCCCAATGAAACAATGCTGGGGTG 1405  
 Db 1261 GGAGCAAACTTAAAGTCTCCCTTGTCTCGGAAAGAGCTCCCAATGAAACAATGCTGGGGTG 1320  
 QY 1406 GTCTGCAACCCCGTGAAGGCGAGAGGCAATGCACTCCGTGAGAGGTCCAGAGGCC 1465  
 Db 1321 GTCTGCAACCCCGTGAAGGCGAGAGGCAATGCACTCCGTGAGAGGTCCAGAGGCC 1380  
 QY 1466 CAGAGAGTCTTGGGGGAGCCCATGGGTGCAATGGGAGGCGGTGAGAGATAGAGAGCT 1525  
 Db 1381 CAGAGAGTCTTGGGGGAGCCCATGGGTGCAATGGGAGGCGGTGAGAGATAGAGAGCT 1440  
 QY 1526 TCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGGAAAGTCCAGC 1585  
 Db 1441 TCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGGAAAGTCCAGC 1500  
 QY 1586 ACCGAGCTGAGACCTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGTA 1645  
 Db 1501 ACCGAGCTGAGAGCTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGTA 1560  
 QY 1646 GAAAGAGTGTCAACAGTGTAAAGGCCCCAAAGCTTAAGGCTTCTCCGCCACCTTTTC 1705  
 Db 1561 GAAAGAGTGTCAACAGTGTAAAGGCCCCAAAGCTTAAGGCTTCTCTCCGCCACCTTTTC 1620  
 QY 1706 CCCCCCGGGTGTGATCGTGGCAGAGTGTGCCCCGTGTGCAATGCGGGGAGGAAAGAGTG 1765  
 Db 1621 CCCCCCGGGTGTGATCGTGGCAGAGTGTGCCCCGTGTGCAATGCGGGGAGGAAAGAGTG 1680  
 QY 1766 GCTTCAATGATCGTGTGTGCTCCCGTGAATGTTCAAGAGTGTGCGTGTGAGACAGAG 1825  
 Db 1681 GCTTCAATGATCGTGTGTGCTCCCGTGAATGTTCAAGAGTGTGCGTGTGAGACAGAG 1740  
 QY 1826 GTGGCTTCCGTGTGTGCGCGGGGCAATGACCCAGAGTGTGCTTTGTGTGAGAGAACAGAGTG 1885  
 Db 1741 GTGGCTTCCGTGTGTGCGCGGGGCAATGACCCAGAGTGTGCTTTGTGTGAGAGAACAGAGTG 1800  
 QY 1886 GGCCTGTGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGAAAGAGAGAGCGTG 1945  
 Db 1801 GGCCTGTGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGAAAGAGAGAGCGTG 1860  
 QY 1946 GAGGACTTGAGAAATGATTAAGGCGACACCGTTCAGAGAGCGCAGAGATCGGCTTACT 2005  
 Db 1861 GAGGACTTGAGAAATGATTAAGGCGACACCGTTCAGAGAGCGCAGAGATCGGCTTACT 1920  
 QY 2006 AGATGCAAGACCCCGCAGAG 2025  
 Db 1921 AGATGCAAGACCCCGCAGAG 1940

RESULT 9  
 LOCUS CR608666 2083 bp mRNA linear HTC 21-JUL-2004  
 DEFINITION Full-length cDNA clone CS0D6004YL23 of B cells (Ramos cell line) of  
 ACCESSION CR608666 Homo sapiens (human).

VERSION CR608666.1 GI:50489473  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 2083)  
 AUTHORS Li, W.B., Gruber, C., Jesssee, J., and Polyes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue  
 2 (bases 1 to 2083)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submission  
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 JOURNAL - Web : www.genoscope.cns.fr)  
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES Location/Qualifiers  
 source 1..2083  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D6004YL23"  
 /rissue\_type="B cells (Ramos cell line)"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 95.8%; Score 1940; DB 3; Length 2083;  
 Best local Similarity 100.0%; Pred. No. 0;  
 Matches 1940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 GCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGACACCAGGCATATG 145  
 Db 1 GCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGACACCAGGCATATG 60  
 QY 146 GCGAACAAGCTATGGAACCTATGACACGCCCACTGATGCTATACCCAGGCTCAGA 205  
 Db 61 GCGAACAAGCTATGGAACCTATGACACGCCCACTGATGCTATACCCAGGCTCAGA 120  
 QY 206 CCACTGCAACTATGGGAGAGACCGCTTATGCAACTCTTATGAGAGAGCCCTCCACTGGTT 265  
 Db 121 CCACTGCAACTATGGGAGAGACCGCTTATGCAACTCTTATGAGAGAGCCCTCCACTGGTT 180  
 QY 266 ATACTACTCAACTGCCCCCGAGGACATACAGCCAGGCTGTCCAGGGGTATGGCACTGTG 325  
 Db 181 ATACTACTCAACTGCCCCCGAGGACATACAGCCAGGCTGTCCAGGGGTATGGCACTGTG 240  
 QY 326 CTTATGATACCAACCTGCTATACAGTACACCAACCAAGGCTCTTATGAGAGCTCACTG 385  
 Db 241 CTTATGATACCAACCTGCTATACAGTACACCAACCAAGGCTCTTATGAGAGCTCACTG 300  
 QY 386 CATATGGCACTAGCGCTGCTTATCCAGCTTATGGGAGAGAGCGAGCCACTGACCTTA 445  
 Db 301 CATATGGCACTAGCGCTGCTTATCCAGCTTATGGGAGAGAGCGAGCCACTGACCTTA 360  
 QY 446 CAAAGCCGAGATGAAAACAAGCCCACTGAGACTAGTCAACCTCATATGACACAGGGG 505  
 Db 361 CAAAGCCGAGATGAAAACAAGCCCACTGAGACTAGTCAACCTCATATGACACAGGGG 420  
 QY 506 GTTACAACCAAGCCGCTTATGAGATATGACAGAGTAACTACAGTTATCCCAAGTACTG 565  
 Db 421 GTTACAACCAAGCCGCTTATGAGATATGACAGAGTAACTACAGTTATCCCAAGTACTG 480  
 QY 566 GGAGCTACCCCAATGAGCCAGTCACTGACCTCCATCCCTCATACAGTATTCCT 625  
 Db 481 GGAGCTACCCCAATGAGCCAGTCACTGACCTCCATCCCTCATACAGTATTCCT 540

QY	626	CTTACACAGCCGACATAGTTATATGATATGAGCAGTACTCTCTGACGACAACACCTTATGGGCAAC	685
Db	541	CTTACACAGCCGACATAGTTATATGATATGAGCAGTACTCTCTGACGACAACACCTTATGGGCAAC	600
QY	686	CGAGCAGCTATGACACGACGATGACTATGGTCAACAAAGCCTATGGGCAAGCCTC	745
Db	601	CGAGCAGCTATGACACGACGATGACTATGGTCAACAAAGCCTATGGGCAAGCCTC	660
QY	746	CCACTAGTATCCACCCCAATCGGATGCTCTACGACCAAGCTCCAAAGTCATATATGCGCAAC	805
Db	661	CCACTAGTATCCACCCCAATCGGATGCTCTACGACCAAGCTCCAAAGTCATATATGCGCAAC	720
QY	806	AGAGCAGCAGCTACCGGCGACGACGATTCATTCCGACAGAACCAACCCGATGACATGGGTG	865
Db	721	AGAGCAGCAGCTACCGGCGACGACGATTCATTCCGACAGAACCAACCCGATGACATGGGTG	780
QY	866	TTTATGGCGACGAGAGTCTGAGAGATTTTTCGGAACGAGAGAGAACCCGAGCATGAGTGGCC	925
Db	781	TTTATGGCGACGAGAGTCTGAGAGATTTTTCGGAACGAGAGAGAACCCGAGCATGAGTGGCC	840
QY	926	CTGATTAACCGGGGCGACGGGAGAGAGGGGGATTGTGATCGTAGAGGCATGACGAGAGTGGGC	985
Db	841	CTGATTAACCGGGGCGACGGGAGAGAGGGGGATTGTGATCGTAGAGGCATGACGAGAGTGGGC	900
QY	986	GGGGAGAGAGACGCGGTGAAATGGCGACGCGCTGAGAGACGAGGTGGCTTCATTAAGCCTG	1045
Db	901	GGGGAGAGAGACGCGGTGAAATGGCGACGCGCTGAGAGACGAGGTGGCTTCATTAAGCCTG	960
QY	1046	GTGGACCCATGGATGGAAGACCAAGATCTTGTAATCTTAGGCCCACTGTGATGATCAATATGAG	1105
Db	961	GTGGACCCATGGATGGAAGACCAAGATCTTGTAATCTTAGGCCCACTGTGATGATCAATATGAG	1020
QY	1106	ACTCTGACAAACAGTCCAAATTTATGTACAAAGATTTAAATGACAGTGTGACTCTAGATGATC	1165
Db	1021	ACTCTGACAAACAGTCCAAATTTATGTACAAAGATTTAAATGACAGTGTGACTCTAGATGATC	1080
QY	1166	TGGCAGACTTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAAACTGGGCAACCCA	1225
Db	1081	TGGCAGACTTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAAACTGGGCAACCCA	1140
QY	1226	TGATCCACATCTACCTGGACAAAGAAACAGGAAAGCCCAAGAGCGCATGCCACAGTGTCT	1285
Db	1141	TGATCCACATCTACCTGGACAAAGAAACAGGAAAGCCCAAGAGCGCATGCCACAGTGTCT	1200
QY	1286	ATGAAGACCCACCCACCTGCGCAAGCGTCCGTGGAAATGTTGATGGGAAAGATTTTCAAG	1345
Db	1201	ATGAAGACCCACCCACCTGCGCAAGCGTCCGTGGAAATGTTGATGGGAAAGATTTTCAAG	1260
QY	1346	GGAGCAAACTTAAATGCTCCCTGCTCGGAAAGAAAGCCCTCAATGAAACGTAATGGGGGTG	1405
Db	1261	GGAGCAAACTTAAATGCTCCCTGCTCGGAAAGAAAGCCCTCAATGAAACGTAATGGGGGTG	1320
QY	1406	GTCTCCACACCCCGTGAAGGCGAGAGGCATGCCACCACTCCGTGGAGSTCCAGAGGCC	1465
Db	1321	GTCTCCACACCCCGTGAAGGCGAGAGGCATGCCACCACTCCGTGGAGSTCCAGAGGCC	1380
QY	1466	CAGAGAGTCTCTGGGGGACCCCATATGGTTCGATGGGAGGGCGCTGTGAGAGATATGAGAGGCT	1525
Db	1381	CAGAGAGTCTCTGGGGGACCCCATATGGTTCGATGGGAGGGCGCTGTGAGAGATATGAGAGGCT	1440
QY	1526	TCCCTCTCAAGAGAGCCCGGGGTTTCCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCAGC	1585
Db	1441	TCCCTCTCAAGAGAGCCCGGGGTTTCCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCAGC	1500
QY	1586	ACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTTCCGCTGGA	1645
Db	1501	ACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTTCCGCTGGA	1560
QY	1646	GAAACGAGTGAACAACAGTGTAGAGGCCCAAGCCTGAAAGGCTTCTCCCGGCAACCTTTC	1705
Db	1561	GAAACGAGTGAACAACAGTGTAGAGGCCCAAGCCTGAAAGGCTTCTCCCGGCAACCTTTC	1620

QY	1706	GGCCCCGGGGTGGATCGGTGGCAGAGTGGCCCTGGTGGCAATCGGGGAGGAAGAGTG	1765
Db	1621	CGCCCCGGGGTGGATCGGTGGCAGAGTGGCCCTGGTGGCAATCGGGGAGGAAGAGTG	1680
QY	1766	GCCCTCATGGATCGGTGGTGTCCCGGTGGAATGTTTCAGAGGTGGCCGTGTGGAGACAG	1825
Db	1681	GCCTCATGGATCGGTGGTGTCCCGGTGGAATGTTTCAGAGGTGGCCGTGTGGAGACAG	1740
QY	1826	GTGGCTTCCGTGTGGCCCGGGGCATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAGTG	1885
Db	1741	GTGGCTTCCGTGTGGCCCGGGGCATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAGTG	1800
QY	1886	GCCCTGGGGGGCCCCCTGGACCTTTGATGGACAAGTGGGAGGAAGAAGGAGGAGCGTG	1945
Db	1801	GCCCTGGGGGGCCCCCTGGACCTTTGATGGACAAGTGGGAGGAAGAAGGAGGAGCGTG	1860
QY	1946	GAGGACCTGGAAAAATGATTAAGGCGAGCACGTCAGAGCGCAGAGATCGGCCCTTACT	2005
Db	1861	GAGGACCTGGAAAAATGATTAAGGCGAGCACGTCAGAGCGCAGAGATCGGCCCTTACT	1920
QY	2006	AGATGCAAGACCCCGCAGA 2025	
Db	1921	AGATGCAAGACCCCGCAGA 1940	

RESULT 10			
CR604639			
LOCUS	2236 bp	mrna	linear
DEFINITION	full-length cDNA clone C6SD013YN08 of T cells (jurkat cell line)		
	Cot 10-normalized of Homo sapiens (human) .		
			HTC 21-JUL-2004

ACCESSION	CR604639
VERSION	CR604639.1
KEYWORDS	HTC; CNSLT_CDNA.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)

REFERENCE  
1 (pages 1 to 2236)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** Li, W. B., Gruber, C., Jesse, J. and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished

REMARK  
Contact : Feng Liang Email : [lliang@lritech.com](mailto:lliang@lritech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue  
REFERENCE  
2 (bases 1 to 2236)

JOURNAL TITLE  
AUTHORS  
Genoscope.  
Direct Submission  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

COMMENT  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqreg@genoscope.cns.fr](mailto:seqreg@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers were used for the RT-PCR. Note : 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers were used for the RT-PCR. Note :

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

```
FEATURES
source
location/Qualifiers
1. .2236
/organism="Homo sapiens"
```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODJ013YN08"
/tissue_type="T cells (Jurkat cell line) Cot

```

```
10-normalized"
/plasmid="pCMVSPORT_6"
```

Query Match	92.1%	Score 1866;	DB 3;	Length 2236;
Best Local Similarity	99.8%	Pred. No. 0;		

Matches	18/9/	conservative	0/	mismatches	0/	indels	3/	gaps	1/
19	GAGGAGGAGGAGAAATGCGCTCCACGATTA	CAGTACCTATAGCCAACTGACGC	78						

Db  
1 GAGGAGGAAGAGAAAAATGGCGCTCCACGGATTACAGTACCTATAGCCCAAGCTGCAGCG 60

QY 79 CAGAGGGCTTACAGTGTCTTACACCGCCAGCCCACTAAGATATGACAGACCAACCCAG 138  
 DB 61 CAGAGGGCTTACAGTGTCTTACACCGCCAGCCCACTAAGATATGACAGACCAACCCAG 120  
 QY 139 GCATATGGGCAACAAAGCTATGAACTATGAAAGAGCCCACTGATGTCAGTATACCCAG 198  
 DB 121 GCATATGGGCAACAAAGCTATGAACTATGAAAGAGCCCACTGATGTCAGTATACCCAG 180  
 QY 199 GCTCAGACCACTGCACTATGGGCAAGCGCTATGCAACTTCTTATGAGACAGCTCCC 258  
 DB 181 GCTCAGACCACTGCACTATGGGCAAGCGCTATGCAACTTCTTATGAGACAGCTCCC 240  
 QY 259 ACTGGTTATATCTTCCAACTGCCCCCAGGACATACAGCCAGCTGTCAGGGGTATGGC 318  
 DB 241 ACTGGTTATATCTTCCAACTGCCCCCAGGACATACAGCCAGCTGTCAGGGGTATGGC 300  
 QY 319 ACTGGTCTTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGAGCT 378  
 DB 301 ACTGGTCTTATGATACCACTGCTACAGTCAACCAACCCAGGCTCTTATGAGCT 360  
 QY 379 CAGTCTGCAATATGCACTCAGCTGCTTATCCAGCTATGGGCAAGCCAGCCACT 438  
 DB 361 CAGTCTGCAATATGCACTCAGCTGCTTATCCAGCTATGGGCAAGCCAGCCACT 420  
 QY 439 GCACCTAACAGACCCGAGATGGAACAAAGCCCACTGAGACTGATCACTCAATCTAGC 498  
 DB 421 GCACCTAACAGACCCGAGATGGAACAAAGCCCACTGAGACTGATCACTCAATCTAGC 480  
 QY 499 ACAGGGGGTTTCAACACGAGCCAGCTAGATATGAGACAGATTACTACAGTTATCCCCAG 558  
 DB 481 ACAGGGGGTTTCAACACGAGCCAGCTAGATATGAGACAGATTACTACAGTTATCCCCAG 540  
 QY 559 GTACTTGGAGCTACCCCATGACAGCCAGTCACTGCACTTCACTCCCTCCAGC 618  
 DB 541 GTACTTGGAGCTACCCCATGACAGCCAGTCACTGCACTTCACTCCCTCCAGC 600  
 QY 619 TATTTCTCTTACAGACCCGATAGTATGATCAGAGCACTTCTCAGCAACCACTTAT 678  
 DB 601 TATTTCTCTTACAGACCCGATAGTATGATCAGAGCACTTCTCAGCAACCACTTAT 660  
 QY 679 GGGCAACCGAGCACTATGAGACAGAGATAGTATGTCACAAAGCAGCTATGGGCAAG 738  
 DB 661 GGGCAACCGAGCACTATGAGACAGAGATAGTATGTCACAAAGCAGCTATGGGCAAG 720  
 QY 739 CAGCTCTCCCACTAGTTACCAACCCCAACTGATCTTCAACAGCCAGCTCCAGTCAATAT 798  
 DB 721 CAGCTCTCCCACTAGTTACCAACCCCAACTGATCTTCAACAGCCAGCTCCAGTCAATAT 780  
 QY 799 AGCCAACAGAGCAAGCTATGAGACAGAGATAGTATGTCACAGCAACCCAGTATAGC 858  
 DB 781 AGCCAACAGAGCAAGCTATGAGACAGAGATAGTATGTCACAGCAACCCAGTATAGC 840  
 QY 859 ATGGGTGTTTATGGGAGAGAGTCTGGAGATTTTCCGGACACAGAGAGAACCCGAGCATG 918  
 DB 841 ATGGGTGTTTATGGGAGAGAGTCTGGAGATTTTCCGGACACAGAGAGAACCCGAGCATG 900  
 QY 919 AGTGGCTCTGATTAACCGGGGCAAGGGGAGATTTGATCGTGGAGCATGAGCAGA 978  
 DB 901 AGTGGCTCTGATTAACCGGGGCAAGGGGAGATTTGATCGTGGAGCATGAGCAGA 960  
 QY 979 GGTGGGCGGGAGAGAGACGGCGGTGAAATGGGACGGCTTGGAGAGCGAGGTGGCTTCAAT 1038  
 DB 961 GGTGGGCGGGAGAGAGACGGCGGTGAAATGGGACGGCTTGGAGAGCGAGGTGGCTTCAAT 1017  
 QY 1039 AAGCTCTGTGAGACCCATGATGAAGACAGATCTTATGATAGGCCCACTGTATGATCA 1098  
 DB 1018 AAGCTCTGTGAGACCCATGATGAAGACAGATCTTATGATAGGCCCACTGTATGATCA 1077  
 QY 1099 GATGAAGACTCTGACAAACAGTGAATTTATGTAACAAGATTAATGACAGTGTACTCTTA 1158  
 DB 1078 GATGAAGACTCTGACAAACAGTGAATTTATGTAACAAGATTAATGACAGTGTACTCTTA 1137

QY 1159 GATGATCTGGACAGACTTCTTAAAGCAGTGGGGTGTTAAGATGAACAAGAACTGGG 1218  
 DB 1138 GATGATCTGGACAGACTTCTTAAAGCAGTGGGGTGTTAAGATGAACAAGAACTGGG 1197  
 QY 1219 CAACCCATGATCAACATCTACTGAGCAAGAAACAGAGAAAGCCCAAGGCGATGCCACA 1278  
 DB 1198 CAACCCATGATCAACATCTACTGAGCAAGAAACAGAGAAAGCCCAAGGCGATGCCACA 1257  
 QY 1279 GTGTCTTATGAAGACCAACCACTGCCAAGGCTGCCGTGAAATGTTTGAATGGAAAGAT 1338  
 DB 1258 GTGTCTTATGAAGACCAACCACTGCCAAGGCTGCCGTGAAATGTTTGAATGGAAAGAT 1317  
 QY 1339 TTTCAAGGAGGCAAACTTAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAACATATG 1398  
 DB 1318 TTTCAAGGAGGCAAACTTAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAACATATG 1377  
 QY 1399 CGGGGTGTCTGACACCCGAGGAGGAGGAGATGACACCACTCCGTTGAGAGTCCA 1458  
 DB 1378 CGGGGTGTCTGACACCCGAGGAGGAGGAGATGACACCACTCCGTTGAGAGTCCA 1437  
 QY 1459 GAGGCCCAAGAGGTCCTGGGAGGACCCATGAGGTGCAATGAGAGGCGTGGAGAGATAGA 1518  
 DB 1438 GAGGCCCAAGAGGTCCTGGGAGGACCCATGAGGTGCAATGAGAGGCGTGGAGAGATAGA 1497  
 QY 1519 GAGGGCTTCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTTGGAGAGGAAAC 1578  
 DB 1498 GAGGGCTTCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTTGGAGAGGAAAC 1557  
 QY 1579 GTCCAGACCCAGCTGAGAGCTGAGCAGTGTCCCAATCCGGGTTGTGGAACCAAGATTC 1638  
 DB 1558 GTCCAGACCCAGCTGAGAGCTGAGCAGTGTCCCAATCCGGGTTGTGGAACCAAGATTC 1617  
 QY 1639 GCCTGAGAAACAGAGTGCAACAGTGTAAAGCCCAAGCCTGAAGGCTTCTCCGCCA 1698  
 DB 1618 GCCTGAGAAACAGAGTGCAACAGTGTAAAGCCCAAGCCTGAAGGCTTCTCCGCCA 1677  
 QY 1699 CCTTTTCCGCCCCCGGGTGTATCTGTGCAAGGTGGCCCTGTGGCATGCCGGAGGA 1758  
 DB 1678 CCTTTTCCGCCCCCGGGTGTATCTGTGCAAGGTGGCCCTGTGGCATGCCGGGGAGGA 1737  
 QY 1759 AGAGGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1818  
 DB 1738 AGAGGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1797  
 QY 1819 GACAGAGTGTGCTTCCGTGTGTGCGGGGATGAGCCGAGGTGCTTGTGTGAGAGAGA 1878  
 DB 1798 GACAGAGTGTGCTTCCGTGTGTGCGGGGATGAGCCGAGGTGCTTGTGTGAGAGAGA 1857  
 QY 1879 CGAGGTGGCTTGGGGGGCCCC 1900  
 DB 1858 CGAGGTGGCTTGGGGGGCCCC 1879

RESULT 11  
 AK034755  
 LOCUS 2373 bp mRNA linear HTC 03-APR-2004  
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region  
 and neck cDNA, RIKEN full-length enriched library, clone:943003M11  
 product:Ewing sarcoma homolog, full insert sequence.  
 AK034755.1 GI:26084174  
 ACCESSION AK034755.1  
 VERSION 1  
 KEYWORDS HTC; GAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL 99279253  
 MEDLINE 10349636  
 PUBMED  
 REFERENCE 2



**AUTHORS**  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

**TITLE**  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

**JOURNAL**  
Genome Res. 10 (10), 1617-1630 (2000)

**PUBLISHED**  
20493974  
11042159

**REFERENCE**

**AUTHORS**  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Niishi, K., Kitamura, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Niehlme, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Iwaw, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kitz, A. and Hayashizaki, Y.

**TITLE**  
RIKEN integrated sequence analysis (RISA) system-384-format

**JOURNAL**  
RIKEN Integrated sequence analysis (RISA) system-384-format

**MEDLINE**  
Genome Res. 10 (11), 1757-1771 (2000)

**PUBLISHED**  
20530913  
11076681

**REFERENCE**

**AUTHORS**  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

**TITLE**  
Functional annotation of a full-length mouse cDNA collection

**JOURNAL**  
Nature 409, 685-690 (2001)

**PUBLISHED**  
11076681

**REFERENCE**

**AUTHORS**  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**TITLE**  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL**  
Nature 420, 563-573 (2002)

**PUBLISHED**  
6 (bases 1 to 2373)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizemoto, K., Hirooka, T., Hirozane, T., Horii, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Niishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazune, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sobabe, Y., Tagami, M., Tagawa, A., Takehashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsutsui-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-redesign@riken.jp, Fax: 81-45-503-9216]

**PUBLISHED**  
Pax: 81-45-503-9216

**COMMENT**  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

**FEATURES**  
**SOURCE**  
1..2373  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:943003IM11"  
/db\_xref="taxon:10090"  
/clone="943003IM11"  
/tissue\_type="embryonic body between diaphragm region and neck"

**misc\_feature**  
/clone\_jib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
43..2011

Query Match	Best Local Similarity	87.4%	Score 1769.8	DB 3	Length 2373
Matches 1873	Conservative	0	Mismatches 132	Indels 4	Gaps 2
polya_signal	putative	2358.2363			
polya_site	/note="putative"	2373			
ORIGIN	/note="putative"				
Query Match	Best Local Similarity	87.4%	Score 1769.8	DB 3	Length 2373
Matches 1873	Conservative	0	Mismatches 132	Indels 4	Gaps 2
polya_signal	putative	2358.2363			
polya_site	/note="putative"	2373			
ORIGIN	/note="putative"				



Qy 900 AGAGAGAACCGGAGCATGAGTGGCCCTGATTAACCGGGGCGAGGGGAAAGGGGGATTGGA 959  
 Db 907 AGAGAGAACCGGAGCTTGAGTGGCCCTGATTAACCGGGGCGAGGGGAAAGGGGGATTGGA 966  
 Qy 960 TCGTGAGGAGCATGAGAGAGTGGCGGGGAGAGAGACGCGGTGAAATGGGACGCTGG 1019  
 Db 967 TCGTGAGGAGCATGAGAGAGTGGCGGGGAGAGAGACGCGGTGACTGG---CGCGTGG 1023  
 Qy 1020 AGAGCGAGGTGGCTTCAATTAAGCCTGTGTGACCCATGTATGAGAGACCATCTTGAATCT 1079  
 Db 1024 AGAGCGAGGTGGCTTCAATTAAGCCTGTGTGACCCATGTATGAGAGACCATCTTGAATCT 1083  
 Qy 1080 AGGCGCATCTGTATGATCCAGATGAAGAATCTGACAAAGGCAATTTATGTACAAAGATTT 1139  
 Db 1084 AGGCGCATCTGTATGATCCAGATGAAGAATCTGACAAAGGCAATTTATGTACAAAGATTT 1143  
 Qy 1140 AATGACAGTGTGACTGTAGTATGATCTGTGACAGATCTTCTTTAAGCAGTGTGGGTTGTTAA 1199  
 Db 1144 AATGACAGTGTGACTGTAGTATGATCTGTGACAGATCTTCTTTAAGCAGTGTGGGTTGTTCA 1203  
 Qy 1200 GATGACAAAGAACTGGGCAACCATGATTCACATTTACTGTGACAAAGAAACAGAA 1259  
 Db 1204 GATGACAAAGAACTGGGCAACCATGATTCATATCTACTGTGATGAAGAGACAGAA 1263  
 Qy 1260 GCCCAAGGGGATGCCACAGTGTCTATGAAGACCCACATGCCCAAGGCTGGCTGTGA 1319  
 Db 1264 GCCCAAGGGGATGCCACAGTGTCTATGAAGATCCACAACTGCAAAAGCTGGCGGTGA 1323  
 Qy 1320 ATGTTTGAATGAGAAATTTTCAAGGAGAGCAAACTTAAAGTCTCCCTGTGCGGAAGAA 1379  
 Db 1324 ATGTTTGAATGAGAAATTTTCAAGGAGAGCAAACTTAAAGTCTCTTCCCGCAAAAGAA 1383  
 Qy 1380 GCTTCCATGACATGATGCGGGGTGTCTGCCACCCCGTGAAGGCGAGGATGCCACC 1439  
 Db 1384 GCTTCCATGACATGATGCGGGGTGTCTGCCACCCCGTGAAGGCGAGGATGCCACC 1443  
 Qy 1440 ACCACTCGTGAAGGTCCAGAGGCGCGAGAGGTCTCGGGGAGACCCATGTGTGCGCAATGG 1499  
 Db 1444 ACCACTCGTGAAGGTCTCGTGGCCAGAGGCGCTGAGAGAGCCATGTGTGCGCAATGG 1503  
 Qy 1500 AGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGAGACCCCGGGGTTCCCGAGGGAA 1559  
 Db 1504 AGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGAGAGCCCGGGGTTCCCGAGGGAA 1563  
 Qy 1560 CCCCTCTGAGAGAGAAACGTCCAGACCCAGAGTGAAGATGTGTCCCAATCCGAGG 1619  
 Db 1564 CCCCTCTGAGAGAGAAATGTCCAGACCCAGAGTGAAGATGTGTCCCAATCCGAGG 1623  
 Qy 1620 TTGTGAAACCAAGAACTTCCCTGAGAGAAAGATGCAACAGTGAAGGCCCCCAAGGCC 1679  
 Db 1624 TTGTGAAACCAAGAACTTCCCTGAGAGAAAGATGCAACAGTGAAGGCCCCCAAGGCC 1683  
 Qy 1680 TGAAGGCTTCTCCCGGACACCTTTCCGCCCCCGGGGTGTGATGTGTGAGAGAGTGGCCC 1739  
 Db 1684 TGAAGGCTTCTCCCGGACACCTTTCCGCCCCCGGGGTGTGATGTGTGAGAGAGTGGCCC 1743  
 Qy 1740 TGGTGTGATCGGGGAGAAAGAGTGTCTCATGTATGTGTGTGTGTGTGTGTGTGTGT 1799  
 Db 1744 TGGTGTGATCGGGGAGAAAGAGTGTCTCATGTATGTGTGTGTGTGTGTGTGTGTGT 1803  
 Qy 1800 CAAAGGT 1859  
 Db 1804 CAAAGGT 1863  
 Qy 1860 TGGCTTGT 1919  
 Db 1864 TGGCTTGT 1923  
 Qy 1920 GATGGAGAGAAAG 1979  
 Db 1924 GATGGAGAGAAAG 1983  
 Qy 1980 TCAGAGAGCGAGAGATGCGCCCTACTAGGA 2008

Db 1984 TCAGAGAGCGAGAGATGCGCCCTACTAGGA 2012

RESULT 12  
 AK019460  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AK019460 2107 bp mRNA linear HTC 03-APR-2004  
 Mus musculus 18 days pregnant adult female placenta and extra  
 embryonic tissue cDNA, RIKEN full-length enriched library,  
 clone:3830417B11 product:Ewing sarcoma homolog, full insert  
 sequence.  
 AK019460  
 AK019460.1 GI:12859676  
 HTC; CAP trapper.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 TITILE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCES  
 AUTHORS

1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 The PANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2107)

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCES  
 AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

ADACHI, J., AIZAWA, K., AKIHARA, S., AKIMURA, T., ARAI, A., AONO, H.,  
 ARAKAWA, T., BONO, H., CARNINCI, P., FUKUDA, S., FUKUNISHI, Y.,  
 HIRAKAWA, T., HORI, F., IMOCANI, K., ISHII, Y., ITOH, M., IZAWA, M.,  
 KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KOMNO, H., KOUNDA, M.,  
 KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., NISHI, K.,  
 NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y., OKIDO, T., OWA, C.,  
 SAITO, H., SAITO, R., SAKAI, C., SAKAI, S., SANO, H., SASEKI, D.,  
 SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y.,  
 SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TANAKA, T.,  
 TEJIMA, Y., TOYA, T., YAMAMURA, T., YASUNISHI, A., YOSHIDA, K.,  
 YOSHINO, M., YAMAMURA, M. and Hayashizaki, Y.  
 Direct SubMISSION  
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),



QY 1341 TCAAGGAGCAAACTTAAGTCTCCCTTGTCCGAAAGACCTCCATGACAGTATGCG 1400  
 Db 1318 TCAAGGAGCAAACTTAAGTCTCTTGTCCGAAAGACCTCCATGACAGTATGCG 1377  
 QY 1401 GGGTGTCTCCCAACCCCGTGAAGGAGCATGCCACCACTCCGTGAGAGTCCAGG 1460  
 Db 1378 GGGAGGATATCCCACTCTGTGAGGGCAGGGGATGCCACCACTCTGTGAGAGTCTTGG 1437  
 QY 1461 AGGCCCAAGAGTCTTGGGGAGCCCATGGGTGCGATGGAGAGCCGTGAGAGATAGAGG 1520  
 Db 1438 TGGCCCAAGAGGCTCTGAGAGGACCCATGGGTGCGATGGAGAGCCGTGAGAGATAGAGG 1497  
 QY 1521 AGGTTTCCCTCCAAAGAGAGACCCCGGGGTTCGCCAGGAAACCTCTGTGAGAGAGAAAGT 1580  
 Db 1498 GGGCTTCTCTCCAAAGAGAGGCCCCCAAGGCTCCAGAGAGAAACCTCTGTGAGAGAGAAAGT 1557  
 QY 1581 CCAGACCCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGGTTGTGAGAAACCAAGAACTTCGCG 1640  
 Db 1558 CCAGACCCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGGCTGTGAGAAACCAAGAACTTCGCG 1617  
 QY 1641 CTGAGAGACAGAGTGAACCAAGTGTAAAGCCCAAGCTGAAAGGCTTCTCCGCGCAC 1700  
 Db 1618 TTGAGAGACAGAAATGCAACCAAGTGTAAAGCCCAAGCTGAAAGGCTTCTCCGCGCAC 1677  
 QY 1701 CTTTCCGCCCCCGGGTGTATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGGAGAGAG 1760  
 Db 1678 CTTTCCACCTTCGGGGTGTATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGGAGAGAG 1737  
 QY 1761 AGGTGGCTCATGAGATCGTGGTGTCTCCGGTGAATGTTTCAAGAGTGGCCGCTGTGAGAG 1820  
 Db 1738 AGAGAGACTATGAGACCTGTGTGTCTTGGAGAAATTTTCAAGAGTGGCCAGAGTGTGAGA 1797  
 QY 1821 CAGAGGTGGCTTCCTGGTGTGGCCCGGAGCATGACCGAGTGTGCTTTGTGAGAGAGACG 1880  
 Db 1798 CAGAGAGGCTTCGAGAGTGGCCCGTGAATGAGACCGAGTGTGCTTTGTGAGAGAGACG 1857  
 QY 1881 AGGTGGCTTCCTGGTGTGGCCCGGAGCATGACCGAGTGTGCTTTGTGAGAGAGAGG 1940  
 Db 1858 AGGTGGCTTCCTGGTGTGGCCCGTGAATGAGACCGAGTGTGCTTTGTGAGAGAGAGG 1917  
 QY 1941 AGGTGAGAGACTGTAAGAAATGATTAAGGCGAGCAACCTGAGAGGCGAGAGTGGGCC 2000  
 Db 1918 AGGTGAGAGACTGTAAGAAATGATTAAGGCGAGCAACCTGAGAGGCGAGAGTGGGCC 1977  
 QY 2001 CTACTAGA 2008  
 Db 1978 CTACTAGA 1985  
 RESULT 13  
 AK049743  
 LOCUS 2269 bp mRNA linear HTC 03-APR-2004  
 DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length  
 enriched library, clone: C530046A18 product: Bwing sarcoma homolog,  
 full insert sequence.  
 AK049743  
 VERSION AK049743.1 GI:26093614  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Miyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Harada, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL  
 MEDLINE 20530913  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 JOURNAL  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS  
 6  
 (bases 1 to 2269)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
 Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submision  
 JOURNAL  
 MEDLINE 11076861  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-Ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 COMMENT  
 CNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/  
 Location/Qualifiers  
 1..2269  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM:DB:C530046A18"  
 /db\_xref="taxon:10090"  
 /clone="C530046A18"  
 /issue\_type="spinal cord"  
 /clone\_id="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="12 days embryo"  
 48..1903  
 /note="Bwing sarcoma homolog (MGD|MGI:99960, GB|NM\_007968,  
 evidence: BASTV, 99%, match=2172)  
 putative"  
 2251..2256  
 polyA\_signal



RESULT	14
BX325713	
LOCUS	
DEFINITION	BX325713 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CS0D1039YJ24 5-PRIME, mRNA sequence.
ACCESSION	BX325713
VERSION	BX325713.2
KEYWORDS	GI:46280663
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	L1 (bases 1 to 1049)
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Poljansky, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	unpublished (2001)
COMMENT	On May 2, 2003 this sequence version replaced gi:30344479.

For more information about this cluster, see <http://www.genoscope.cns.fr/cdna?c=CS0A5012ZH02QP1&c=10589.f>

```

SOURCE
1. .1049
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1039YJ24"
/tissue_type="P16ACENTA COT 25-NORMALIZED"
/clone_id="Homo sapiens P16ACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

Query Match	48.7%	Score 987	DB 5	Length 1049
Best Local Similarity	98.3%	Pred. No. 1e-238		
Matches 1029	Conservative	8	Mismatches	6
			Indels	4
			Gaps	4

QY	3	CGGACGTTGAGAGAACGAGGAGGAAAGGAGAAATGGCGTCACACGATTTACATGACCTA	62
Db	7	CGGACGTTGAGAGAAACGAGGAGGAGGAGGAAATGGCGTCACACGATTTACATGACCTA	66
QY	63	TAGCCAAAGCTGCAGCGCAGCAGGGCTACAGTGCCTTACACGCGCAGCCACTCAAGATA	122
Db	67	TAGCCAAAGCTGCAGCGCAGGAGGAGGAGGCTACAGTGCCTTACACCGCCAGCCCACTCAAGATA	126
QY	123	TGCAACAGCCACCCAGGCGATATGGGCAACAAGCTATATGAACTCTATGACAGGCCACTGA	182
Db	127	TGCACAGCCACCCAGGCGATATGGGCAACAAGCTATATGAACTCTATGACAGGCCCACTGA	186
QY	183	TGTGAGCTATATCCGAGGCTCAGACCACTGCACACTATATGGGAGACCGGCTATGCACTTC	242
Db	187	TGTGAGCTATATCCGAGGCTCAGACCACTGCACCACTATATGGGAGACCGGCTATGCACTTC	246
QY	243	TTATATGACAGGCTCCCACTGGTTATATCTATCTTCACTGCCCCCGAGGATATCAAGCAGCC	302
Db	247	TTATATGACAGGCTCCCACTGGTTATATCTATCTTCACTGCCCCCGAGGATATCAAGCAGCC	306
QY	303	TGTCACAGGGATATGGCACTGGTGCCTTATGATATACCAACACTGCTACAGTCAACCAACCA	362
Db	307	TGTCACAGGGATATGGCACTGGTGCCTTATGATATACCAACACTGCTACAGTCAACCAACCA	366
QY	363	GGCTCTCTATGCAAGCTCAGTCTGCATATAGCACTTCAAGCTGCTTATTCACGCGTATGGCA	422

Db	367	GGCTCTTATGCACTAGTCTGATATGGCACTCAACCTGTCTTATTCAGCTATGGGCA	426
Qy	423	GGAGCGACGACCTGCACTGACCTTACAGACCGGAGATGMAAACAGCCCACTGAGCTAG	482
Db	427	GCAGCGACGACCTGCACTGACCTTACAAACCGCA - GATGGAAACAAAGCCCACTGAGACTAG	485
Qy	483	TCAACCTCAATCTATGCAAGGGGGTTTCAACGACGCCAGCTGAGATATGGACAGATTA	542
Db	486	TCAACCTCAATCTATGCAAGGGGGTTTCAACGACGCCAGCTGAGATATGGACAGATTA	545
Qy	543	CTACAGTTATCCCGAGGTACTCTGGAGCTACCCCATGACAGCCAGTCACTGCACTTCATC	602
Db	546	CTACAGTTATCCCGAGGTACTCTGGAGGTACCCCATGACAGCTCACTGCACTTCATC	605
Qy	603	CTACCTCTCTTACAGGCTATTCCTCTTACAGCGCGCACTAGTTATGATCAAGCACTTATCTC	662
Db	606	CTACCTCTCTTACAGGCTATTCCTCTTACAGCGCGCACTAGTTATGATCAAGCACTTATCTC	665
Qy	663	TCAGCAGAACCTATATGGGCAACCGAGCAGTATGACAGCAGAGTATGTTGTTCAACA	722
Db	666	TCAGCAGAACCTATATGGGCAACCGAGCAGTATGACAGCAGAGTATGTTGTTCAACA	725
Qy	723	AAGCAGCTATATGGGCAAGAGCTCTCCACTAAGTTACCAACCCCAACCTGATCTTACAGCCA	782
Db	726	AAGCAGCTATATGGGCAAGAGCTCTCCACTAAGTTACCAACCCCAACCTGATCTTACAGCCA	785
Qy	783	AGCTCCAAGTATATATGACCAACGACAGCAGAGCTAGAGGCAAGCAGATTCATTCCGACA	842
Db	786	AGCTCCAAGTATATATGACCAACGACAGCAGAGCTAGAGGCAAGCAGATTCATTCCGACA	845
Qy	843	GGACCACTCCCGATGACATGGGTGTTTATATGGGCAAGAGTCTGAGAGATTTTCCGACCAAG	902
Db	846	-GRCACACCCCGATGACATGGGTGTTTATATGGGCAAGAGTCTGAGAGATTTTCCGACCAAG	904
Qy	903	AGAAACCCGAGCATATGATGGCCCTGATTAACCGGGGAGAGGGGAAAGCGGGATTTGATCG	962
Db	905	AGAAACCCGAGCATATGATGGCCCTGATTAACCGGGGAGAGGGGAAAGCGGGATTTGATCG	964
Qy	963	TGAGAGCATATGACAGAGTGGGCGGGAGAGAGACGCGATGATATGGGACGCTGAGAGA	1022
Db	965	TGAGAGCATATGACAGAGTGGGCGGGAGAGAGACGCGATGATATGGGACGCTGAGAGA	1023
Qy	1023	GGAGGTGCTTCAATAAGCTGGTGG 1049	
Db	1024	SCGAGGTGGCTTC - ATTAAGCTTGGTGG 1049	

RESULT	15
LOCUS	BK353313
DEFINITION	BK353313 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens CDNA clone SCODC004YER22 5-PRIME, Eukarya sequence.
ACCESSION	BK353313
VERSION	BK353313.2
KEYWORDS	EST.
SOURCE	GI:46291060
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1008) L.J.W.B., Gruber,C., Jesse,J. and Polayes,D. Full-length cDNA libraries and normalization unpublished (2001) On May 5, 2003 this sequence version replaced gi:30371752.
AUTHORS	Contact: Genoscope
TITLE	Genoscope - Centre National de Sequencage
JOURNAL	2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
COMMENT	Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-coligodIdT primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen. This sequence belongs to sequence cluster



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:41:37 ; Search time 4248.83 Seconds

(Without alignments)  
11541.209 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012

Sequence: 1 aggaacggatggagaaatgggca.....agcgacagatggcgccctac 1012

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 470823 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: GenBank1.\*

1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_dat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_atg:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1012	100.0	1971	12	BT007796
2	1012	100.0	1988	6	AX714390
3	1012	100.0	1988	5	AK056681
4	1012	100.0	2182	9	BC004817
5	1012	100.0	2390	6	BC0867364
6	1010.4	99.8	2371	6	A36460
7	1010.4	99.8	2371	6	AR080100
8	1010.4	99.8	2390	6	AK411125
9	1010.4	99.8	2390	6	HSEWS
10	1008.8	99.7	2026	6	CQ850483
11	1008.8	99.7	2026	6	AK127624
12	1001	98.9	2326	6	CQ730352
13	996	98.4	2164	9	AK056309
14	996	98.4	2189	9	AK056309
15	996	98.4	2189	9	CR456490
16	996	98.4	2364	9	BC011048
17	993.4	98.2	7293	9	BSM808920
18	955.4	94.4	2473	9	AK026270
19	911.2	90.0	193559	2	AC137500

20	908	89.7	155815	2	AC018774	AC018774 Homo sapi
21	908	89.7	180718	9	AL596087	AL596087 Human DNA
22	908	89.7	182501	2	AC011221	AC011221 Homo sapi
23	895.2	88.5	2189	10	BC068226	BC068226 Mus muscu
24	879.2	86.9	2188	6	AX305537	AX305537 Sequence
25	879.2	86.9	2188	10	MMEMS	X79233 M.musculus
26	879.2	86.9	2396	6	BC083960	BC083960 Xenopus 1
27	860	85.0	1783	6	CQ721057	CQ721057 Sequence
28	817.6	80.8	132906	2	AC109802	AC109802 Canis fam
29	817.6	80.8	163104	2	AC110672	AC110672 Canis fam
30	779	77.0	221444	2	AC113313	AC113313 Mus muscu
31	779	77.0	247757	2	AC121282	AC121282 Mus muscu
32	750	74.1	226400	2	AC128482	AC128482 Rattus no
33	750	74.1	225241	2	AC098231	AC098231 Rattus no
34	750	74.1	263925	2	AC106522	AC106522 Rattus no
35	653.6	64.6	1462	6	CQ434162	CQ434162 Sequence
36	595.6	58.9	2440	5	AJ719366	AJ719366 Gallus ga
37	562.8	55.6	220708	2	AC150903	AC150903 Mus muscu
38	562.8	55.6	249523	2	AC079537	AC079537 Mus muscu
39	559.6	55.3	274656	2	AC079538	AC079538 Mus muscu
40	559.6	55.3	274656	2	AC079538	AC079538 Mus muscu
41	552.8	54.6	257383	2	AC079494	AC079494 Mus muscu
42	524.6	51.8	270337	2	AC099212	AC099212 Rattus no
43	513.8	50.8	110000	2	AC149218_2	Continuation (3 of
44	510.6	50.5	160334	2	AC079549	AC079549 Mus muscu
45	508	50.2	221924	2	AC105889	AC105889 Rattus no

## ALIGNMENTS

RESULT 1	BT007796	1971 bp	mrna	linear	SYN 13-MAY-2003
LOCUS	Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1				
DEFINITION	BT007796	partial cds.			
ACCESSION	BT007796.1	GI:30584430			
VERSION	BT007796.1	GI:30584430			
KEYWORDS	Full cDNA.				
SOURCE	Synthetic construct				
ORGANISM	Synthetic construct				
REFERENCE	1 (bases 1 to 1971)				
AUTHORS	Kalnina,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koudinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.				
TITLE	Cloning of human full-length cDNAs in BD Creator(TM) System Donor vector				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1971)				
AUTHORS	Kalnina,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koudinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.				
TITLE	Direct Substitution				
JOURNAL	Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA				
COMMENT	This cDNA clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each cDNA has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The cDNA has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDONR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: http://bioinfo.clontech.com/orfclones.				
FEATURES	Location/Qualifiers				
Source	1..1971				
	/organism="synthetic construct"				
	/mol_type="mRNA"				
	/db_xref="taxon:32630"				
	/clone="GH00297L1.0"				
	/clone_1b="BD Creator(TM) CDS library derived from MGC				





Db	1040	CATCTACCTGGACAAGAAACGAAAGGCCAAAGCCGATGCGACAGTGTCCTATTGAAGA	1099
Oy	301	CCCAACCCACTGGCAAGGCTGCCGTGGAAATGTTTGAATGGAAAGATTTTCAAGGAGCA	360
Db	1100	CCCAACCCACTGGCAAGGCTGCCGTGGAAATGTTTGAATGGAAAGATTTTCAAGGAGCA	1159
Oy	361	ACTTAAAGTCTCCCTTGCTGGGAAGAAGCTCCATATGAACGTAAGCGGGGGGTCTGCC	420
Db	1160	ACTTAAAGTCTCCCTTGCTGGGAAGAAGCTCCATATGAACGTAAGCGGGGGGTCTGCC	1219
Oy	421	ACCCCGTAGGGCAGAGGCATGCGCAACCACTCCGTGGAGGTCCAGAGGCGCCAGAGAG	480
Db	1220	ACCCCGTAGGGCAGAGGCATGCGCAACCACTCCGTGGAGGTCCAGAGGCGCCAGAGAG	1279
Oy	481	TCCTGGGGAGACCATGATGGGTCCGATGGGAGGCCGTGAGAGAGATAGAGAGGCTTCC	540
Db	1280	TCCTGGGGAGACCATGATGGGTCCGATGGGAGGCCGTGAGAGAGATAGAGAGGCTTCC	1339
Oy	541	AAAGAGACCCCGGGGTTTCCCGAGGGGAACCCCTCTGGAAGAGAAACGTCCAGACCCGAGC	600
Db	1340	AAAGAGACCCCGGGGTTTCCCGAGGGGAACCCCTCTGGAAGAGAAACGTCCAGACCCGAGC	1399
Oy	601	TGAGAGACTGGCAGTGTCCCAATCCGGGTGTGTGAAACAGAACTTCGCTGGAGAAACGA	660
Db	1400	TGAGAGACTGGCAGTGTCCCAATCCGGGTGTGTGAAACAGAACTTCGCTGGAGAAACGA	1459
Oy	661	GTGCAACCAAGTGTAAAGGCCCCAAAGCCTGAAGGCTTCTCCCGCACCCCTTCCGCCCC	720
Db	1460	GTGCAACCAAGTGTAAAGGCCCCAAAGCCTGAAGGCTTCTCCCGCACCCCTTCCGCCCC	1519
Oy	721	GGGTGGTAAATCGTGGCAGAGGTGGCCCTGTGTGGCATTCGGGGGAGGAAGAAGTGGCTCAT	780
Db	1520	GGGTGGTAAATCGTGGCAGAGGTGGCCCTGTGTGGCATTCGGGGGAGGAAGAAGTGGCTCAT	1579
Oy	781	GGATCGTGTGGTCCCGGTGGAATGTTTCAAGAGTGAACCGGTGGAGACAGAGGTGGCTT	840
Db	1580	GGATCGTGTGGTCCCGGTGGAATGTTTCAAGAGTGAACCGGTGGAGACAGAGGTGGCTT	1639
Oy	841	CCGTGTGTGACCGGGGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	900
Db	1640	CCGTGTGTGACCGGGGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	1699
Oy	901	GGGGCCCCCTTGGACCTTGTATGAAACAGATGGAGGAAGAAAGAGAGGACGTGAGAGACC	960
Db	1700	GGGGCCCCCTTGGACCTTGTATGAAACAGATGGAGGAAGAAAGAGAGGACGTGAGAGACC	1759
Oy	961	TGGAATAATGATTAAGGCGGACACCGTTCAGAGAGGCGAGAAATCGGCCCTTAC	1012
Db	1760	TGGAATAATGATTAAGGCGGACACCGTTCAGAGAGGCGAGAAATCGGCCCTTAC	1811

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AK056681		1986 bp mRNA linear PRI 30-JAN-2004	AK056681						
		Homo sapiens cDNA FLJ32119 f1s, clone FBLML1000034, highly similar to RNA-BINDING PROTEIN EMS.	AK056681						
		AK056681.1 GI:16552152							
		oligo capping, f1s (full insert sequence).							
		Homo sapiens (human)							
		Homo sapiens							
		Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
1		Ota, T., Suzuki, Y., Nishikawa, T., Otsubi, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, I., Isono, Y., Nakamura, Y., Nagahashi, K., Murakami, K., Yasuda, T., Iwainaga, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Futaya, T., Tokawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Takikawa, M., Yamazaki, T., Nishimura, K., Ishibashi, T., Yamashita, H.,							

JOURNAL PUBMED REFERENCE AUTHORS	TITLE
2	<p>Murakawa, K., Fujimori, K., Tanaai, H., Kinata, M., Watanabe, M., Hiraoka, S., Chida, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houcha, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Muesahino, K., Yunki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigami, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hama, H., Watanabe, M., Komatsu, T., Mzushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuhio, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs</p> <p>Nat. Genet. 36 (1), 40-45 (2004)</p> <p>14702039</p>
	<p>Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,</p>

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
NEDO human cDNA sequencing project	Unpublished	3 (bases 1 to 1986)	Isooga,T., Otsuki,T. and Sugiyama,T.	Direct Submission		
Submitted (24-OCT-2001)						
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0012, Japan						
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)						
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing:						
Research Association for Biotechnology (RAB) ; cDNA library						
construction: Helix Research Institute (HRI) (supported by Japan						
Key Technology Center etc.); 5'- & 3'- and one pass sequencing: RAB,						
HRI, and Biotechnology Center, National Institute of Technology and						
Evaluation; clone selection for full insert sequencing: RAB and						
HRI.						

```

FEATURES
    source
        1..1988
            location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="PEBL100034"
                /cell_type="peripheral blood mononuclear cells (HPBMC593)"
                /clone_1lb="PEBL1"
                /note="Cloning vector: pME18SFL3-primary culture, peripheral blood mononuclear cells"
            12..1814
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="BAB71252.1"
                /db_xref="GI:16552153"
                /translation="MASTDYSTYSQAQAQGSAYTAQPTQGYAQTQAGVAGGGSYGTY
                GPDVAYTOAQTATYAGTAAVATSYGQPTGTTPTTAPQASQPVQAGVAGADTTT
                AYTQDASAAQASVGTQBPAPVAYGQQAATPTSYSTQPTSVDQSSYQNTQGG
                PSSYQDQGSYGQOQSSSYGQAPPTSPYPTGYSQAQPSQYISQOQSSSYGQSSFPQDHPSS
                MGATYQGSAGSGSGEGRKRSYSGFPNRRGRGSGFDRGMSRGKRGQGGNGSKDEERG
                FNRFGQPMDEGPDLDLGPVPDPEDSDNSAIYVQGLNDSYVLLDLADLFFQCGGVKRN
                KRQGGPMIHTLYLDETKCPGDDATVSYEDPPTKAAAEWEDGQFQSSKTKVSLARKK

```



```

/db_xref="taxon:9606"
/clone="MGC:5432 IMAGE:3449145"
/class_type="Placenta, choriocarcinoma"
/clone_id="NH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
1. .2182
/gene="EWSR1"
/note="synonym: EMS"
/db_xref="LOCUSID:2130"
/db_xref="MIM:133450"
20. .1990
/gene="EWSR1"
/codon_start=1
/product="Ewing sarcoma breakpoint region 1, isoform EMS"
/db_xref="GI:13435963"
/db_xref="LOCUSID:2130"
/db_xref="MIM:133450"
/translation="MSTDYVTSQAAAOQGSAYTAAPOTGVAGTQTAAYGQSGVGY
GQPTDVAQQTATATGCGATVTSYGQPTGTTPATPAASQPVQSGVAGADITTTT
ATVTTTQASAAASGATGYPAYVQGGQALATPAPRQDGNKPTTSGQSSSTGGNO
PSLQVGSNSTVQVPEVSMQYATPSPYPTPTTSISTQTPSYSDSSTSQNTGQPS
SYGQSSVYGQSSVGGQPPSTPPTQVGSYQASPSIQSSSYGQSSFRDHPHSMGCS
KPGGPMDEGGFDLGPVPEBDSNSALYVQGLNDSVTLIDLFPKQCVVQNNKRR
TGCGMIYLDEKGEKPGATVSEYDPTAKAAVEWFPKQFQSGKVLARKKPP
NMNSRGLIYDREKGMKPTPLRGPGGPGGDMBMGGRGSGGPPRGPSRGRN
PSGGCGVQHAAGMCCPNRCGQNNAPMTTECNCKAPPEBFTLPPPPPCGCRGCG
PGGCGKRGGLGNDGPGGGMFGGCGRGRGRGRGRGRGRGFGGGRGGPGGPDP
MEWGGRRGGRGCGMGMDGHRGRRRRPP"

```

## ORIGIN

Query Match	100.0%	Score 1012;	DB 9;	Length 2182;
Best Local Similarity	100.0%	Pred. No. 6.8e-197;		
Matches 1012;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AGACGCGGTGAATGGCAGCGCTGGAGAGCGAGTGGCTTCAATAAAGCTGGTGGAC	60	
Dp	976	AGGACGCGGTGAATGGCAGCGCTGGAGAGCGAGTGGCTTCAATAAAGCTGGTGGAC	1035	
Qy	61	CATGATTAAGAACCAATCTTGGATCTTGGCCACCTGTAAATCAGATGAAGACTCTGA	120	
Dp	1036	CATGATTAAGAACCAAGATCTTGGATCTTGGCCACCTGTAAATCAGATGAAGACTCTGA	1095	
Qy	121	CAACAGTCAATTTATGTACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	180	
Dp	1096	CAACAGTCAATTTATGTACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	1155	
Qy	181	CTTCTTTAAGCAGTGTGGGTGTGTTAAGATGACAAGAACTGGGCAACCATGATCCA	240	
Dp	1156	CTTCTTTAAGCAGTGTGGGTGTGTTAAGATGACAAGAACTGGGCAACCATGATCCA	1215	
Qy	241	CATCTAAGCTTGGACAAAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAAGA	300	
Dp	1216	CATCTAAGCTTGGACAAAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAAGA	1275	
Qy	301	CCCAACCACTGCGCAAGGCTGCCCGTGGATATGTTTGAATGGAAAGATTTTCAAGGAGCA	360	
Dp	1276	CCCAACCACTGCGCAAGGCTGCCCGTGGATATGTTTGAATGGAAAGATTTTCAAGGAGCA	1335	
Qy	361	ACTTAAAGTCTCCCTTGTCTCGGAAGAACTCTCAATGAACAGTATGCGGGGTGGTCTGCC	420	
Dp	1336	ACTTAAAGTCTCCCTTGTCTCGGAAGAACTCTCAATGAACAGTATGCGGGGTGGTCTGCC	1395	
Qy	421	ACCCGTGAGGGCAGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGGCCCCAGAGAG	480	
Dp	1396	ACCCGTGAGGGCAGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGGCCCCAGAGAG	1455	
Qy	481	TTCCTGGGGGAGCCCAATGGGTGCGCATGGGAGGCGCTGGAGGAATATGAGAGGCTTCCCTCC	540	
Dp	1456	TTCCTGGGGGAGCCCAATGGGTGCGCATGGGAGGCGCTGGAGGAATATGAGAGGCTTCCCTCC	1515	

QY	541	AAGAGAACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAGAAACGTCAGACACGAGC	600
Db	1516	AAGAGAACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAGAAACGTCAGACACGAGC	1575
QY	601	TGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGACTTGCCTGAGAGACAGA	660
Db	1576	TGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGACTTGCCTGAGAGACAGA	1635
QY	661	GTGCAACCAAGTAAAGGCCCAAGACCTGAAGGCTTCCTCCCGCACCCCTTCCGGCCCC	720
Db	1636	GTGCAACCAAGTAAAGGCCCAAGACCTGAAGGCTTCCTCCCGCACCCCTTCCGGCCCC	1695
QY	721	GCGTGTGATGTGTGGCAGAGGTGGCCCTGTGTGCATGCGGGAGGAAGAGGTGGCTCAT	780
Db	1696	GCGTGTGATGTGTGGCAGAGGTGGCCCTGTGTGCATGCGGGAGGAAGAGGTGGCTCAT	1755
QY	781	GGATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGGCTT	840
Db	1756	GGATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGGCTT	1815
QY	841	CCGTGTGTGGCCGGGGCATGACCCGAGGTGGCTTTGGTGAAGAAAGACGAGGTGGCCCTGG	900
Db	1816	CCGTGTGTGGCCGGGGCATGACCCGAGGTGGCTTTGGTGAAGAAAGACGAGGTGGCCCTGG	1875
QY	901	GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGCC	960
Db	1876	GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGCC	1935
QY	961	TGAAAAAATGATTAAGCGCAGACCGTCAAGAGCGCAGAGATGGCGCTTAC	1012
Db	1936	TGAAAAAATGATTAAGCGCAGACCGTCAAGAGCGCAGAGATGGCGCTTAC	1987

RESULT 5				
LOCUS LOC67364	CQ867364	2390 bp	DNA	linear PAT 13-SEP-2004
DEFINITION	Sequence 1 from Patent EP1455190.			
ACCESSION	CQ867364			
VERSION	CQ867364.1	GI:51997589		
KEYWORDS	.			
SOURCE ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 Obendorf M. and Wolf S.			
AUTHORS	Modulation of the activity of nuclear receptors via EMS			
TITLE	Patent: EP 1455190-A 1 08-SEP-2004;			
JOURNAL	Schering Aktiengesellschaft (DE)			
FEATURES	location/Qualifiers			
source	1..2390			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
	44..2014			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAH33891.1"			
	/db_xref="GI:51997590"			
	/translation="MASTDYSTVYSQAAGQGSAYTAOTPGYAOTTQAVGQQSGYTV			
	GQPDVSVTAOQTATYGCTAATSYSCQPTGYTTPAPAYSQPVGGYGTADVTITTT			
	ATVTTTQASAAAGSYGTOPAYVGGOOPATATRRDQGNKPFETSOPOSTGYNQ			
	PSLGGDSNYSPDVPPGVSYPMOPTAPRASPPTSYSSTPOYSQSSISQONTYGGS			
	SYGQSSSYGQSSSYGQOPTSPYSPQTSYGAQBQYQOQSSYQOSSFDFDHSSMG			
	VYGESGSGFSFGPGENRSMSGDDNNGRGGRDRGMSRGSGRGGMGSGKGRGFRN			
	KPGQMDGEPDLIDGPVEDDEDSDNSAIIVQGLINDSVTLDDLADFQKCGVVMMNR			
	TQGQMIIHYLDKETKGKDAIVSYEDPFAKAWEWFDGKDFQGSXLYSLAKKEPP			
	MNSRRGCLPREGRGMPPLRGSGGSGGWGRMRGRGDGGRFPGRSGRSGRSG			
	PSGGCQNHQARQAMQCNPQGCNNQFMARTTECNCKAXPKDEGFLPPFPPGGRGG			
	GPGMRGRGRGLNDRGSGPMFRGGRGRGDGSGFFGSGMGRGRGGRGRGPGPPOP			
	LMEQGRGRGRGSGPMGRGGRGHEHQERRDPR"			
ORIGIN				

Query Match 100.0%; Score 1012; DB 6; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-197;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGGACCGGATGGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 60  
 1000 AGGACCGGATGGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 1059

61 CATGATGAAAGACCAAGATCTTATCTAGGCCCCCCTGTAGATCCAGATGAACCTCTGA 120  
 1060 CATGATGAAAGACCAAGATCTTATCTAGGCCCCCCTGTAGATCCAGATGAACCTCTGA 1119

121 CAACAGTGAATTTATGTAAGAGATTAATGACAGTGTACTATGATGATCTGGAGA 180  
 1120 CAACAGTGAATTTATGTAAGAGATTAATGACAGTGTACTATGATGATCTGGAGA 1179

181 CTTCTTTAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGGCAACCATGATCCA 240  
 1180 CTTCTTTAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGGCAACCATGATCCA 1239

241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCCATGTCTTATGAGA 300  
 1240 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCCATGTCTTATGAGA 1299

301 CCCACCACTGCAAGGCTGCGGTGAGATGGTGTGAGGAAAGATTTTCAAGGGAGCAA 360  
 1300 CCCACCACTGCAAGGCTGCGGTGAGATGGTGTGAGGAAAGATTTTCAAGGGAGCAA 1359

361 ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTCCAGTAAACAGTATGGGGGTGTCTGCC 420  
 1360 ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTCCAGTAAACAGTATGGGGGTGTCTGCC 1419

421 ACCCTGGAAGGACAGAGATGCAACCACTCTCGGTGAGTCTCAGAGAGCCCAAGAG 480  
 1420 ACCCTGGAAGGACAGAGATGCAACCACTCTCGGTGAGTCTCAGAGAGCCCAAGAG 1479

481 TCCCTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 540  
 1480 TCCCTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 1539

541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 600  
 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 1599

601 TGGAGATGCGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGAGAGAAAGA 660  
 1600 TGGAGATGCGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGAGAGAAAGA 1659

661 GTGCAACCAAGTGAAGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTCCGCCCCC 720  
 1660 GTGCAACCAAGTGAAGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTCCGCCCCC 1719

721 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGGCGGAGAGAAAGGTGGCTCAT 780  
 1720 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGGCGGAGAGAAAGGTGGCTCAT 1779

781 GGATCTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGGAACAAGAGTGGCTT 840  
 1780 GGATCTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGGAACAAGAGTGGCTT 1839

841 CCGTGTGTGCGCGGGACATGACCGAGGTGCTTTGTGTGAGAGAGACGAGGTGGCCCTGG 900  
 1840 CCGTGTGTGCGCGGGACATGACCGAGGTGCTTTGTGTGAGAGAGACGAGGTGGCCCTGG 1899

901 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGGAGAAC 960  
 1900 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGGAGAAC 1959

961 TGGAAAAATGATTAAGCGGACACCTTCAGAGACGAGAGATGGCCCTTAC 1012  
 1960 TGGAAAAATGATTAAGCGGACACCTTCAGAGACGAGAGATGGCCCTTAC 2011

RESULT 6  
 A36460  
 LOCUS 2371 bp DNA linear PAT 05-MAR-1997  
 DEFINITION Sequence 1 from Patent WO9323549.  
 ACCESSION A36460  
 VERSION A36460.1 GI:2293778  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 1 (bases 1 to 2371)  
 Aurias, A., Delattre, O., Desmaziere, C., Melot, T., Peter, M.,  
 Plougastel, B., Thomas, G. and Zucman, J.  
 NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
 RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE  
 DEVELOPMENT OF CANCEROUS TUMORS  
 Patent: WO 9323549-A 1 25-NOV-1993;  
 CENTRE NAT RECH SCIENT (FR)  
 Other publication FR 2691475 931126  
 Other publication JP 8500964T 960206.  
 COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1. 2371  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="FOETUS"

Query Match 99.8%; Score 1010.4; DB 6; Length 2371;  
 Best Local Similarity 99.9%; Pred. No. 1.4e-196;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGGACCGGATGGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 60  
 981 AGGACCGGATGGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 1040

61 CATGATGAAAGACCAAGATCTTATCTAGGCCCCCCTGTAGATCCAGATGAACCTCTGA 120  
 1041 CATGATGAAAGACCAAGATCTTATCTAGGCCCCCCTGTAGATCCAGATGAACCTCTGA 1100

121 CAACAGTGAATTTATGTAAGAGATTAATGACAGTGTACTATGATGATCTGGAGA 180  
 1101 CAACAGTGAATTTATGTAAGAGATTAATGACAGTGTACTATGATGATCTGGAGA 1160

181 CTTCTTTAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGGCAACCATGATCCA 240  
 1161 CTTCTTTAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGGCAACCATGATCCA 1220

241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCCATGTCTTATGAGA 300  
 1221 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCCATGTCTTATGAGA 1280

301 CCCACCACTGCAAGGCTGCGGTGAGATGGTGTGAGGAAAGATTTTCAAGGGAGCAA 360  
 1281 CCCACCACTGCAAGGCTGCGGTGAGATGGTGTGAGGAAAGATTTTCAAGGGAGCAA 1340

361 ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTCCAGTAAACAGTATGGGGGTGTCTGCC 420  
 1341 ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTCCAGTAAACAGTATGGGGGTGTCTGCC 1400

421 ACCCGTGAAGGACAGAGATGCAACCACTCCGTGAGAGTCCAGAGAGCCAGAGAG 480  
 1401 ACCCGTGAAGGACAGAGATGCAACCACTCCGTGAGAGTCCAGAGAGCCAGAGAG 1460

481 TCCCTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGAGATGAGAGAGCTTCCCTCC 540  
 1461 TCCCTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGAGATGAGAGAGCTTCCCTCC 1520

541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAGAAAGCTCCAGACCGAGC 600  
 1521 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAGAAAGCTCCAGACCGAGC 1580

QY 601 TGGAGACTGGCAGTGTCCCATCCGGGTGTGAAAACAGAACTTGGCTGGAGAACAGA 660  
|  
|  
|  
Db 1581 TGGAGACTGGCAGTGTCCCATCCGGGTGTGAAAACAGAACTTGGCTGGAGAACAGA 1640  
|  
|  
|  
QY 661 GTGCAACCAAGTGTAAAGGCCCAAGCCTGAAAGCTTCTCCGCCACCTTTCCGCCCCC 720  
|  
|  
|  
Db 1641 GTGCAACCAAGTGTAAAGGCCCAAGCCTGAAAGCTTCTCCGCCACCTTTCCGCCCCC 1700  
|  
|  
|  
QY 721 GGGTGTGATCTGTGCGAGAGTGGCTCTGTGTCATCCGGGAGAGAAAGAGTGGCTCAT 780  
|  
|  
|  
Db 1701 GGGTGTGATCTGTGCGAGAGTGGCTCTGTGTCATCCGGGAGAGAAAGAGTGGCTCAT 1760  
|  
|  
|  
QY 781 GGATCGTGTGTGTCCTGGGTGGAATGTCAGAGTGGCCGTGTGTGGAACAGAGTGGCTT 840  
|  
|  
|  
Db 1761 GGATCGTGTGTGTCCTGGGTGGAATGTCAGAGTGGCCGTGTGTGGAACAGAGTGGCTT 1820  
|  
|  
|  
QY 841 CCGTGTGTGCGCGGGGACATGACCCAGGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTTG 900  
|  
|  
|  
Db 1821 CCGTGTGTGCGCGGGGACATGACCCAGGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTTG 1880  
|  
|  
|  
QY 901 GGGGCCCCCTGTGACCTTTGATGAAAGATGGAGAAAGAGAGAGAGAGAGAGAGAGAG 960  
|  
|  
|  
Db 1881 GGGGCCCCCTGTGACCTTTGATGAAAGATGGAGAAAGAGAGAGAGAGAGAGAGAGAG 1940  
|  
|  
|  
QY 961 TGGAAAAATGATTAAGGCGAGCAGCTCAAGAGCGAGAGATGGCCCTTAC 1012  
|  
|  
|  
Db 1941 TGGAAAAATGATTAAGGCGAGCAGCTCAAGAGCGAGAGATGGCCCTTAC 1992  
|  
|  
|

RESULT 7  
AR080100 2371 bp DNA linear PAT 31-AUG-2000  
LOCUS  
DEFINITION Sequence 1 from patent US 5968734.  
ACCESSION AR080100  
VERSION AR080100.1 GI:10006835  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2371)  
Aurias, A., Delattre, O., Desmazes, C., Meiot, T., Peter, M.,  
Ploonsastrel, B., Thomas, G. and Zucman, J.  
Nucleic acid corresponding to a gene of chromosome 22 involved in  
recurrent chromosomal translocations associated with the  
development of cancerous tumors, and nucleic acids of fusion  
resulting from said translocations  
Patent: US 5968734-A 1 19-OCT-1999;  
Location/Qualifiers  
1..2371  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Query Match 99.8%; Score 1010.4; DB 6; Length 2371;  
Best Local Similarity 99.9%; Pred. No. 1,4e-196;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGACGGCGTGAATGGGACGCGTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 60  
|  
|  
|  
Db 981 AGGACGGCGTGAATGGGACGCGTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 1040  
|  
|  
|  
QY 61 CATGATGAAGAGACAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
|  
|  
|  
Db 1041 CATGATGAAGAGACAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 1100  
|  
|  
|  
QY 121 CAACAGTGAATTAATGATCAAGATTAATGACAGTGAATTAATGATGATGATGATGATGAT 180  
|  
|  
|  
Db 1101 CAACAGTGAATTAATGATCAAGATTAATGACAGTGAATTAATGATGATGATGATGATGAT 1160  
|  
|  
|  
QY 181 CTTCTTTAGCAGTGTGGGTTGTTAGATGAACAGAGAACTGGGCAACCCATGATCCA 240  
|  
|  
|  
Db 1161 CTTCTTTAGCAGTGTGGGTTGTTAGATGAACAGAGAACTGGGCAACCCATGATCCA 1220  
|  
|  
|

QY 241 CATCTACTGGACAGAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 300  
|  
|  
|  
Db 1221 CATCTACTGGACAGAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 1280  
|  
|  
|  
QY 301 CCACACCACTGCCAAGGCTCCGTGATGATGATGATGATGATGATGATGATGATGATGAT 360  
|  
|  
|  
Db 1281 CCACACCACTGCCAAGGCTCCGTGATGATGATGATGATGATGATGATGATGATGATGAT 1340  
|  
|  
|  
QY 361 ACTTAAAGTCTCCCTGCTCGAAGAAAGCTCCATGAACATGATGCGGGGTGTCTGCC 420  
|  
|  
|  
Db 1341 ACTTAAAGTCTCCCTGCTCGAAGAAAGCTCCATGAACATGATGCGGGGTGTCTGCC 1400  
|  
|  
|  
QY 421 ACCCGTGAAGGAGAGAGAGATGCCACCACTCCGTGAGATCCAGAGGCCAGAGAG 480  
|  
|  
|  
Db 1401 ACCCGTGAAGGAGAGAGAGATGCCACCACTCCGTGAGATCCAGAGGCCAGAGAG 1460  
|  
|  
|  
QY 481 TCCTGGGGGACCAATGAGTGTGATGAGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
|  
|  
|  
Db 1461 TCCTGGGGGACCAATGAGTGTGATGAGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 1520  
|  
|  
|  
QY 541 AAGAGACCCCGGGGTTCCCGAGAGAAACCTCTGTGAGAGAGAAAGCTCCAGACCGAGC 600  
|  
|  
|  
Db 1521 AAGAGACCCCGGGGTTCCCGAGAGAAACCTCTGTGAGAGAGAAAGCTCCAGACCGAGC 1580  
|  
|  
|  
QY 601 TGGAGACTGGCAGTGTCCCATCCGGGTGTGAAAACAGAACTTGGCTGGAGAACAGA 660  
|  
|  
|  
Db 1581 TGGAGACTGGCAGTGTCCCATCCGGGTGTGAAAACAGAACTTGGCTGGAGAACAGA 1640  
|  
|  
|  
QY 661 GTGCAACCAAGTGTAAAGGCCCAAGCCTGAAAGCTTCTCCGCCACCTTTCCGCCCCC 720  
|  
|  
|  
Db 1641 GTGCAACCAAGTGTAAAGGCCCAAGCCTGAAAGCTTCTCCGCCACCTTTCCGCCCCC 1700  
|  
|  
|  
QY 721 GGGTGTGATCTGTGCGAGAGTGGCTCTGTGTCATCCGGGAGAGAAAGAGTGGCTCAT 780  
|  
|  
|  
Db 1701 GGGTGTGATCTGTGCGAGAGTGGCTCTGTGTCATCCGGGAGAGAAAGAGTGGCTCAT 1760  
|  
|  
|  
QY 781 GGATCGTGTGTGTCCTGGGTGGAATGTCAGAGTGGCCGTGTGTGGAACAGAGTGGCTT 840  
|  
|  
|  
Db 1761 GGATCGTGTGTGTCCTGGGTGGAATGTCAGAGTGGCCGTGTGTGGAACAGAGTGGCTT 1820  
|  
|  
|  
QY 841 CCGTGTGTGCGCGGGGACATGACCCAGGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTTG 900  
|  
|  
|  
Db 1821 CCGTGTGTGCGCGGGGACATGACCCAGGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTTG 1880  
|  
|  
|  
QY 901 GGGGCCCCCTGTGACCTTTGATGAAAGATGGAGAAAGAGAGAGAGAGAGAGAGAGAG 960  
|  
|  
|  
Db 1881 GGGGCCCCCTGTGACCTTTGATGAAAGATGGAGAAAGAGAGAGAGAGAGAGAGAGAG 1940  
|  
|  
|  
QY 961 TGGAAAAATGATTAAGGCGAGCAGCTCAAGAGCGAGAGATGGCCCTTAC 1012  
|  
|  
|  
Db 1941 TGGAAAAATGATTAAGGCGAGCAGCTCAAGAGCGAGAGATGGCCCTTAC 1992  
|  
|  
|

RESULT 8  
AX411125 2390 bp DNA linear PAT 14-JUN-2002  
LOCUS  
DEFINITION Sequence 3772 from Patent WO0229103.  
ACCESSION AX411125  
VERSION AX411125.1 GI:21443830  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
Gene expression profiles in liver cancer  
Patent: WO 0229103-A 3772 11-APR-2002;  
Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.  
TITLE  
JOURNAL  
GENE LOGIC INC (US)  
FEATURES  
1..2390  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"





## ORIGIN

Query Match 99.8%; Score 1010.4; DB 9; Length 2390; Best Local Similarity 99.9%; Pred. No. 1,4e-196; Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGGACGGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 60  
 DB 1000 AGGACGGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 1059

QY 61 CATGATGAAGACACAGATCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACTTGA 120  
 DB 1060 CATGATGAAGACACAGATCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACTTGA 1119

QY 121 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGACA 180  
 DB 1120 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGACA 1179

QY 181 CTTCTTTAAGCAGTGGGGTGTAAAGTGAACAAGAACTGGGGACACCCATGATCCA 240  
 DB 1180 CTTCTTTAAGCAGTGGGGTGTAAAGTGAACAAGAACTGGGGACACCCATGATCCA 1239

QY 241 CATCTACCTGACACAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATGAGA 300  
 DB 1240 CATCTACCTGACACAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATGAGA 1299

QY 301 CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 360  
 DB 1300 CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 1359

QY 361 ACTTAAAGTCTCCCTTGTGGAAGAGCCCTCAATGAAGATGAGGGGTGTCTGCC 420  
 DB 1360 ACTTAAAGTCTCCCTTGTGGAAGAGCCCTCAATGAAGATGAGGGGTGTCTGCC 1419

QY 421 ACCCGTGAAGGACAGAGCATGCGCACTCCGTGAGGTCCAGAGGCCACAGAGG 480  
 DB 1420 ACCCGTGAAGGACAGAGCATGCGCACTCCGTGAGGTCCAGAGGCCACAGAGG 1479

QY 481 TCCGTGGGGACCCATGGGTGCGATGGAGGCCGTGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1480 TCCGTGGGGACCCATGGGTGCGATGGAGGCCGTGAGAGATGAGAGGCTTCCCTCC 1539

QY 541 AAGAGGACCCCGGGGTCCCGAGGGGAAACCCCTCGAGAGAGAAACCTCCAGACCGAGC 600  
 DB 1540 AAGAGGACCCCGGGGTCCCGAGGGGAAACCCCTCGAGAGAGAAACCTCCAGACCGAGC 1599

QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGACTTCCCTGAGAAACA 660  
 DB 1600 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGACTTCCCTGAGAAACA 1659

QY 661 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAAGCTTCCCTCCGCAACCTTTCGCCCC 720  
 DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAAGCCTTCCCTCCGCAACCTTTCGCCCC 1719

QY 721 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGAGATGCCGGGAGAAAGAGTGGCTCAT 780  
 DB 1720 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGAGATGCCGGGAGAAAGAGTGGCTCAT 1779

QY 781 GATTCGTGTGTCCCGGTGAATGTTCAGAGGTGCGGTGTGAGAGAACAAGGTGGCTT 840  
 DB 1780 GATTCGTGTGTCCCGGTGAATGTTCAGAGGTGCGGTGTGAGAGAACAAGGTGGCTT 1839

QY 841 CCGTGTGGCCGGGGACATGAGCCGAGGTGCTTGTGAGAGAAACAGAGTGGCTTCC 900  
 DB 1840 CCGTGTGGCCGGGGACATGAGCCGAGGTGCTTGTGAGAGAAACAGAGTGGCTTCC 1899

QY 901 GGGGGCCCTTGGACCTTTGATGAGAAACAGATGGAGAAAGAGAGAGAGTGGAGACC 960  
 DB 1900 GGGGGCCCTTGGACCTTTGATGAGAAACAGATGGAGAAAGAGAGAGAGTGGAGACC 1959

QY 961 TGGAAAAATGATTAAGGCGAGCACCTCTAGAGCGCAGAGATCGGCTTAC 1012  
 DB 1960 TGGAAAAATGATTAAGGCGAGCACCTCTAGAGCGCAGAGATCGGCTTAC 2011

## ORIGIN

Query Match 99.7%; Score 1008.8; DB 6; Length 2026; Best Local Similarity 99.8%; Pred. No. 3.1e-196; Matches 1010; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AGGACCGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 60  
 DB 838 AGGACCGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 897

QY 61 CATGATGAAGACACAGATCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACTTGA 120  
 DB 898 CATGATGAAGACACAGATCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACTTGA 957

QY 121 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGACA 180  
 DB 958 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGACA 1017

QY 181 CTTCTTTAAGCAGTGGGGTGTAAAGTGAACAAGAACTGGGGACACCATGATCCA 240  
 DB 1018 CTTCTTTAAGCAGTGGGGTGTAAAGTGAACAAGAACTGGGGACACCATGATCCA 1077

QY 241 CATCTACCTGACACAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATGAGA 300  
 DB 1078 CATCTACCTGACACAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATGAGA 1137

QY 301 CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 360  
 DB 1138 CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 1197

QY 361 ACTTAAAGTCTCCCTTGTGGAAGAGCCCTCAATGAAGATGAGGGGTGTCTGCC 420  
 DB 1198 ACTTAAAGTCTCCCTTGTGGAAGAGCCCTCAATGAAGATGAGGGGTGTCTGCC 1257

QY 421 ACCCGTGAAGGACAGAGCATGCGCACTCCGTGAGGTCCAGAGGCCACAGAGG 480  
 DB 1258 ACCCGTGAAGGACAGAGCATGCGCACTCCGTGAGGTCCAGAGGCCACAGAGG 1317

QY 481 TCCGTGGGGACCCAGTGGGTGCGATGGAGGCCGTGAGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1318 TCCGTGGGGACCCAGTGGGTGCGATGGAGGCCGTGAGAGAGATGAGAGGCTTCCCTCC 1377

QY 541 AAGAGGACCCCGGGGTCCCGAGGGGAAACCCCTTGTGAGAGAGAAAGTCCAGACCGAGC 600  
 DB 1378 AAGAGGACCCCGGGGTCCCGAGGGGAAACCCCTTGTGAGAGAGAAAGTCCAGACCGAGC 1437

QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGACTTCCCTGAGAAACA 660  
 DB 1438 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGACTTCCCTGAGAAACA 1497

## RESULT 10

COB50483

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 2026

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

Location/Qualifiers

Research Association for Biotechnology (JP)

Patent: EP 147413-A 952 18-AUG-2004;

Full-length human CDNA

Otsuki, T., Wakamatsu, A., Ishii, S., Nagai, K. and Irie, R.

Isogai, T., Yamamoto, J., Nishikawa, T., Isono, Y., Sugiyama, T.,

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eulestomi;

Homo sapiens (human)

COB50483.1 GI:51508695

Sequence 952 from Patent EP147413.

2026 bp. DNA linear PAT 23-AUG-2004

COB50483

Query Match	99.7%	Score 1008.8	DB 9	Length 2026
Best Local Similarity	99.8%	Pred. No. 3.1e-196		
Matches 1010	Conservative	0	Mismatches 2	Indels 0
			Gaps	0
QY 1	AGGACGGGTGGATGAGGCGCTGGAGACGAGGTGGCTTCATTAAGCTGTGTGACC	60		
Db 838	AGGACGGGTGGATGAGGCGCTGGAGACGAGGTGGCTTCATTAAGCTGTGTGACC	897		
QY 61	CATGATGAAAGACCAAGTCTTGATCTTAGGCCCACTGTAGATCCAGATGAAGCTCTGA	120		
Db 898	CATGATGAAAGACCAAGTCTTGATCTTAGGCCCACTGTAGATCCAGATGAAGCTCTGA	957		
QY 121	CAAGAGTGAATTTATGTACAAGATTTAAATGACAGGTGACCTTAGATGTCGGGAGA	180		
Db 958	CAAGAGTGAATTTATGTACAAGATTTAAATGACAGGTGACCTTAGATGTCGGGAGA	1017		
QY 181	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGACCCCATGATCCA	240		
Db 1018	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGACCCCATGATCCA	1077		
QY 241	CATTTACCTTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA	300		
Db 1078	CATTTACCTTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA	1137		
QY 301	CCGACCCACTGCCAAGGCTGCGGTGAATGTTTGAATGGGAAAGATTTTCAAGGGACAA	360		
Db 1138	CCGACCCACTGCCAAGGCTGCGGTGAATGTTTGAATGGGAAAGATTTTCAAGGGACAA	1197		
QY 361	ACTTAAAGTCTCCCTTGCTCGGAAGAAAGCTCCATGAACAGTATGCGGGGTGTCTGCC	420		
Db 1198	ACTTAAAGTCTCCCTTGCTCGGAAGAAAGCTCCATGAACAGTATGCGGGGTGTCTGCC	1257		
QY 421	ACCCCTGAGGGGCAAGAGCATGCCACCACTCCGTGGAGTTCACAAGGCCCAGGAGG	480		
Db 1258	ACCCCTGAGGGGCAAGAGCATGCCACCACTCCGTGGAGTTCACAAGGCCCAGGAGG	1317		
QY 481	TCTGTGGGGGACCCCATGTGGGTGCAATGTGGAGGCGGTGAGAGATGTAGAAGGCTTCCCTCC	540		
Db 1318	TCTGTGGGGGACCCCATGTGGGTGCAATGTGGAGGCGGTGAGAGATGTAGAAGGCTTCCCTCC	1377		
QY 541	AAGAGACCCCGGGGTTCCCGAGGGAAACCCCTGTGAGAGAGAAAGCTCCAGACCGAGC	600		
Db 1378	AAGAGACCCCGGGGTTCCCGAGGGAAACCCCTGTGAGAGAGAAAGCTCCAGACCGAGC	1437		
QY 601	TGGAAGTGGCAGGTGCCAATCCGGGTGTGGAACACAGAACTTGCGCTGGAGAACAGA	660		
Db 1438	TGGAAGTGGCAGGTGCCAATCCGGGTGTGGAACACAGAACTTGCGCTGGAGAACAGA	1497		
QY 661	GTGCAACCAAGTGTAAAGCCCAAGACCTGTGAAGGCTTCTCCCGCACCTTTCCGCCCCC	720		
Db 1498	GTGCAACCAAGTGTAAAGCCCAAGACCTGTGAAGGCTTCTCCCGCACCTTTCCGCCCCC	1557		
QY 721	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCCGGGAGAGAAAGGTGGCTCAT	780		
Db 1558	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCCGGGAGAGAAAGGTGGCTCAT	1617		
QY 781	GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCCGTGTGTGAACAAGAGGTGGCTT	840		
Db 1618	GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCCGTGTGTGAACAAGAGGTGGCTT	1677		
QY 841	CCGTGTGTGGCCGGGGGACATGACCGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCCCTGG	900		
Db 1678	CCGTGTGTGGCCGGGGGACATGACCGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCCCTGG	1737		
QY 901	GGGGCCCCCTTGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGCTGTAGAGACC	960		
Db 1738	GGGGCCCCCTTGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGCTGTAGAGACC	1797		
QY 961	TGGAAGAAATGGAATTAAGGCGAGCACCGTCAAGAGCGAGAGATGGGCTCTTAC	1012		

Db 1798 TGGAAAAATGATTAAGCGGACGCTCAGAGCGGAGAGATGCGCCCTAC 1849

RESULT 12  
LOCUS C0730352 2326 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 16286 from Patent WO2068579.  
ACCESSION C0730352  
VERSION C0730352.1 GI:42303963  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kites, such as nucleic acid arrays, comprising a majority of  
humaneoxons or transcripts, for detecting expression and other uses  
thereof  
JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;  
PB Corporation (NY) (US)  
FEATURES  
source Location/Qualifiers  
1. .2326  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 98.9%; Score 1001; DB 6; Length 2326;  
Best Local Similarity 99.9%; Pred. No. 1.2e-194;  
Matches 1012; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AGGACCGCGGTGGAATGGGCGCGCTGGAAGCGAGGTGCTTCAATAGCTTGAGACC 60  
DB 952 AGGACCGCGGTGGAATGGGCGCGCTGGAAGCGAGGTGCTTCAATAGCTTGAGACC 1011  
QY 61 CATGATGAAAGGACCAAGATCTTGTATGAGGCCACCTGTAGATCCAGATGAACCTCTGA 120  
DB 1012 CATGATGAAAGGACCAAGATCTTGTATGAGGCCACCTGTAGATCCAGATGAACCTCTGA 1071  
QY 121 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTACTCTAGATGATCTGGCAGA 180  
DB 1072 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTACTCTAGATGATCTGGCAGA 1131  
QY 181 CTTCCTTTAAGCAGTGTGGGCTTTGAATGATGAACAAGAACTGGGCAACCCATGATCCA 240  
DB 1132 CTTCCTTTAAGCAGTGTGGGCTTTGAATGATGAACAAGAACTGGGCAACCCATGATCCA 1191  
QY 241 CATCTACCTGGAAGAAGGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 300  
DB 1192 CATCTACCTGGAAGAAGGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 1251  
QY 301 CCCACCACTGCGCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTTCAAGGAGCA 360  
DB 1252 CCCACCACTGCGCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTTCAAGGAGCA 1311  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTTCAATGAACAGTATGCGGGGTCTTCC 420  
DB 1312 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTTCAATGAACAGTATGCGGGGTCTTCC 1371  
QY 421 ACCCGGTGAGGAGGAGGATGCGACCACTCCGTGAGGTCCAGAGGCCCAAGAGG 480  
DB 1372 ACCCGGTGAGGAGGAGGATGCGACCACTCCGTGAGGTCCAGAGGCCCAAGAGG 1431  
QY 481 TCTCGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCC 540  
DB 1432 TCTCGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCC 1491  
QY 541 AAGAGAGACCCCGGGGTTCCCGAGAGAAACCTCTTGAAGAGAGAAAGCTCCAGACCGAGC 600  
DB 1492 AAGAGAGACCCCGGGGTTCCCGAGAGAAACCTCTTGAAGAGAGAAAGCTCCAGACCGAGC 1551  
QY 601 TGGAGACTGCACTGTCCCAATCCGGGTTGTGAAACCAAGACTTCCCTGAGAAACAGA 660

Db 1552 TGGAGACTGCACTGTCCCAATCCGGGTTGTGAAACCAAGACTTCCCTGAGAAACAGA 1611

QY 661 GTGCAACCACTGTAAAGGCCCCCAAGCTGAAAGCTTCTCCCGCACCTTTCGCCGCC 720  
DB 1612 GTGCAACCACTGTAAAGGCCCCCAAGCTGAAAGCTTCTCCCGCACCTTTCGCCGCC 1671  
QY 721 GGGTGTGATGTCGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGAGTGGCTTCAT 780  
DB 1672 GGGTGTGATGTCGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGAGTGGCTTCAT 1731  
QY 781 GGATCGTGTGTGTCGCCGTGAAATGTTTCAAGAGTGGCCGTGAGAGACAGAGTGGCTT 840  
DB 1732 GGATCGTGTGTGTCGCCGTGAAATGTTTCAAGAGTGGCCGTGAGAGACAGAGTGGCTT 1791  
QY 841 -CCGTGTGTGTCGCCGTGAGAGTGGCTTGTGAGAGAGAGAGAGTGGCTTGTG 899  
DB 1792 CCGTGTGTGTCGCCGTGAGAGTGGCTTGTGAGAGAGAGAGAGTGGCTTGTG 1851  
QY 900 GGGGGGCCCTTGTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGAGAGAGAG 959  
DB 1852 GGGGGGCCCTTGTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGAGAGAGAG 1911  
QY 960 CTGAAAAAATGATTAAGCGGACCGCTCAGAGGCGAGAGATGCGCCCTAC 1012  
DB 1912 CTGAAAAAATGATTAAGCGGACCGCTCAGAGGCGAGAGATGCGCCCTAC 1964

RESULT 13  
LOCUS BC072442 2164 bp mRNA linear PRI 30-JUN-2004  
DEFINITION Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
EWS, mRNA (CDNA clone MGC:87992 IMAGE:6047805), complete cds.  
ACCESSION BC072442  
VERSION BC072442.1 GI:48734726  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heib, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Ruhl, S.W.,  
Villalob, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Boutard, G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,  
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Small, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
TITLE Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 2164)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
CONTACT: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:31:32 / Search time 490.802 Seconds  
(without alignments)  
12206.093 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012

Sequence: 1 aggcagcggtgagatcgga.....agcgagagatcgccctac 1012

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1012	100.0	1988	10	ADA53506	ADA53506 Human cod
2	1012	100.0	2176	6	AAS62262	AAS62262 cDNA sequ
3	1012	100.0	2372	13	ADO86032	ADG86032 Human tum
4	1012	100.0	2390	13	ADS16277	ADS16277 Human CDN
5	1010.4	99.8	2371	2	AAQ50643	AAQ50643 Human Ews
6	1010.4	99.8	2388	5	AAS70647	AAS70647 DNA encod
7	1010.4	99.8	2390	6	ABK84628	ABK84628 Human CDN
8	1010.4	99.8	2390	6	ABK64822	ABK64822 Human ben
9	1010.4	99.8	2390	6	ABN97274	ABN97274 Gene #377
10	1010.4	99.8	2390	13	ACN40903	ACN40903 Tumour-as
11	1008.8	99.7	2026	13	ADR07446	ADR07446 Full leng
12	954.4	94.3	1807	13	ADP56333	ADP56333 Human PRO
13	908	89.7	2177	13	ACN37537	ACN37537 Tumour-as
14	882.4	86.2	2273	6	AAS62623	AAS62623 cDNA sequ
15	879.2	86.9	2188	6	AB193383	AB193383 Mouse 18c
16	653.6	56.3	1462	4	AAL26725	AAL26725 Human bre
17	570	53.2	2299	11	ACN88697	ACN88697 Breast ca
18	538	50.2	3309	2	AAQ50646	AAQ50646 Human Ews
19	508.4	50.2	350	9	ACH40478	ACH40478 Human foe
20	437.2	43.2	568	9	ACH22345	ACH22345 Human adu

## ALIGNMENTS

RESULT 1	ADA53506	standard; cDNA; 1988 BP.
ID	ADA53506	
AC	ADA53506	
XX	20-NOV-2003	(first entry)
DT	XX	
XX	Human coding sequence, SEQ ID 1074.	
DE	XX	
XX	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;	
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;	
KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.	
OS	Homo sapiens.	
XX	PN	EP1293569-A2.
XX	PD	19-MAR-2003.
XX	PF	21-MAR-2002; 2002EP-0006586.
XX	PR	14-SEP-2001; 2001JP-00328381.
PR	24-JAN-2002; 2002US-0350435P.	
PA	(HELI-) HELIX RES INST.	
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.	
XX	Iogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;	
PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamehika I;	
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;	
XX	WPI; 2003-395539/38.	
DR	P-PSDB; ADA55145.	
XX	New polynucleotides encoding full-length polypeptides, e.g. secretory	
PT	and/or membrane proteins, useful for developing medicines for diseases in	
PT	which the gene is involved, or as target molecules for gene therapy.	
XX	claim 1; SEQ ID NO 1074; 205bp; English.	
PS	The present invention relates to novel human secretory or membrane	
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA5433-	

21	419.2	41.4	471	2 AAV6051	AAV6051 EST clone
22	309	30.5	545	3 AAC78275	AAC78275 Human can
23	304.6	30.1	665	13 ADOS4544	ADOS4544 Novel can
24	269.8	26.7	361	8 ABX38686	ABX38686 Bovine ES
25	258.4	25.5	503	4 AAI18034	AAI18034 Probe #79
26	258.4	25.5	503	4 ABA63001	ABA63001 Human foe
27	258.4	25.5	503	4 AAI43027	AAI43027 Probe #11
28	258.4	25.5	503	4 ABA30262	ABA30262 Probe #87
29	258.4	25.5	503	4 AAK37201	AAK37201 Human bon
30	258.4	25.5	503	4 AAK11415	AAK11415 Human bra
31	258.4	25.5	503	4 ABA36871	ABA36871 Human liv
32	258.4	25.5	503	6 ABA11189	ABA11189 Human gen
33	258.4	25.5	503	12 ACH71268	ACH71268 Human gen
34	254.2	25.1	262	12 ACH84968	ACH84968 Human gen
35	253	25.0	253	4 AAI27213	AAI27213 Probe #17
36	253	25.0	253	4 ABA75494	ABA75494 Human foe
37	253	25.0	253	4 AAI56067	AAI56067 Probe #24
38	253	25.0	253	4 ABA40118	ABA40118 Probe #18
39	253	25.0	253	4 AAK50129	AAK50129 Human bon
40	253	25.0	253	4 AAK24064	AAK24064 Human bra
41	253	25.0	253	4 ABA49761	ABA49761 Human liv
42	253	25.0	253	6 ABA33605	ABA33605 Murine gen
43	212.4	21.0	457	4 AAK53923	AAK53923 Murine tr
44	176.8	17.5	1832	12 ADI26112	ADI26112 Human CDN
45	176.6	17.5	684	13 ACN37539	ACN37539 Tumour-as



AD54072). The coding sequences are useful in the gene therapy of diseases caused by abnormalities of the proteins, e.g. cancer, inflammatory diseases, osteoporosis or neurological disease.

Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 10; Length 1988;  
Best Local Similarity 100.0%; Pred. No. 2.9e-233; Indels 0; Gaps 0;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGACGCGGTGAATGAGCGCGTGAAGAGCGAGTGCCTTCAATAGCTGAGTACC 60
DB AGGACGCGGTGAATGAGCGCGTGAAGAGCGAGTGCCTTCAATAGCTGAGTACC 859
QY 61 CATGATGAAGAGCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120
DB CATGATGAAGAGCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 919
QY 121 CAACAGTGCATTTATATTAACAAGATTAAATGACAGTGAATCTAGATGATCTGGGAGA 180
DB CAACAGTGCATTTATATTAACAAGATTAAATGACAGTGAATCTAGATGATCTGGGAGA 979
QY 181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCAATGATCA 240
DB CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCAATGATCA 1039
QY 241 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGATGCCAAGTGTCTATGAAGA 300
DB CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGATGCCAAGTGTCTATGAAGA 1099
QY 301 CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAGGAGCA 360
DB CCCACCACTGCGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAGGAGCA 1159
QY 1100 CCAACCACTGCGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAGGAGCA 1159
DB CCAACCACTGCGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAGGAGCA 1159
QY 361 ACTTAAAGTCTCCCTTCTCGAAGAAAGCTTCAATGAACATGATGCGGGTGTCTGCC 420
DB ACTTAAAGTCTCCCTTCTCGAAGAAAGCTTCAATGAACATGATGCGGGTGTCTGCC 1219
QY 421 ACCCGGTGAGGGGAGAGGATGCCAACAATCTCGTGAAGTCCAGAGGCCCAAGAGG 480
DB ACCCGGTGAGGGGAGAGGATGCCAACAATCTCGTGAAGTCCAGAGGCCCAAGAGG 1220
QY 1220 ACCCGGTGAGGGGAGAGGATGCCAACAATCTCGTGAAGTCCAGAGGCCCAAGAGG 1279
DB ACCCGGTGAGGGGAGAGGATGCCAACAATCTCGTGAAGTCCAGAGGCCCAAGAGG 1279
QY 481 TCCGTGGGGGAGCCCAATGAGGTGCGATGGAGGCGGTGAAGAGATGAGAGGCTTCCCTCC 540
DB TCCGTGGGGGAGCCCAATGAGGTGCGATGGAGGCGGTGAAGAGATGAGAGGCTTCCCTCC 1339
QY 1280 TCCGTGGGGGAGCCCAATGAGGTGCGATGGAGGCGGTGAAGAGATGAGAGGCTTCCCTCC 1339
DB TCCGTGGGGGAGCCCAATGAGGTGCGATGGAGGCGGTGAAGAGATGAGAGGCTTCCCTCC 1339
QY 541 AAGAGAGCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 600
DB AAGAGAGCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 1340
QY 1340 AAGAGAGCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 1399
DB AAGAGAGCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 1399
QY 601 TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGAGAGACAGA 660
DB TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGAGAGACAGA 1400
QY 1400 TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGAGAGACAGA 1459
DB TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGAGAGACAGA 1459
QY 661 GTGAACCAAGTGAAGGCGCCCAAGAGCTTCCCTCCCGCACCTTCCGCCCCC 720
DB GTGAACCAAGTGAAGGCGCCCAAGAGCTTCCCTCCCGCACCTTCCGCCCCC 1460
QY 1460 GTGAACCAAGTGAAGGCGCCCAAGAGCTTCCCTCCCGCACCTTCCGCCCCC 1519
DB GTGAACCAAGTGAAGGCGCCCAAGAGCTTCCCTCCCGCACCTTCCGCCCCC 1519
QY 721 GGGTGTGATCTGTGCGAGAGGTGCGCTGTGTGAGTGCAGGAGAGAAAGAGTGGCTCAT 780
DB GGGTGTGATCTGTGCGAGAGGTGCGCTGTGTGAGTGCAGGAGAGAAAGAGTGGCTCAT 1520
QY 1520 GGGTGTGATCTGTGCGAGAGGTGCGCTGTGTGAGTGCAGGAGAGAAAGAGTGGCTCAT 1579
DB GGGTGTGATCTGTGCGAGAGGTGCGCTGTGTGAGTGCAGGAGAGAAAGAGTGGCTCAT 1579
QY 781 GGAATGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 840
DB GGAATGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1580
QY 1580 GGAATGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1639
DB GGAATGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1639
QY 841 CCGGTGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 900
DB CCGGTGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1640
QY 1640 CCGGTGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1659
DB CCGGTGTGTGTGTCCCGGTGAATGTTCAAGGTGCGCGTGTGAGAGACAGAGTGGCTT 1659
QY 901 GGGGCCCCCTTGAATCTTGTATGAACAGATGGAGAGAAAGAGAGACTGTGAGAGACC 960
DB GGGGCCCCCTTGAATCTTGTATGAACAGATGGAGAGAAAGAGAGACTGTGAGAGACC 1700
QY 1700 GGGGCCCCCTTGAATCTTGTATGAACAGATGGAGAGAAAGAGAGACTGTGAGAGACC 1759
DB GGGGCCCCCTTGAATCTTGTATGAACAGATGGAGAGAAAGAGAGACTGTGAGAGACC 1759

```

QY 961 TGGAAATATGATTAAGGAGACCGTCAAGAGGCGCAGAGATCGGCCCTTAC 1012  
DB TGGAAATATGATTAAGGAGACCGTCAAGAGGCGCAGAGATCGGCCCTTAC 1811

RESULT 2  
AAS62262/C  
ID AAS62262 standard; cDNA; 2176 BP.  
XX  
XX AAS62262;  
AC  
AC  
DT 14-FEB-2002 (first entry)  
XX

DE cDNA sequence #49 encoding novel human secreted protein.

XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
XX immune deficiency disorder; blood disorder; inflammatory disorder;  
XX infectious disorder; gene therapy; antimicrobial; hepatotropic;  
XX immunosuppressive; antineumatic; ss.  
XX  
XX Homo sapiens.  
OS  
XX

XX MO200177291-A2.  
XX  
XX 18-OCT-2001.

XX 29-MAR-2001; 2001MO-US010485.  
XX  
XX 06-APR-2000; 2000US-0195604P.

XX (GENY) GENETICS INST INC.  
XX  
XX

XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
XX Galukota K, Graham JR;  
XX  
XX WPI; 2002-010900/01.

XX New polynucleotides encoding secreted proteins useful for treating e.g.  
XX asthma, HIV and Crohn's disease.  
XX  
XX

XX Claim 1; Page 100; 391bp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
XX which encode human secreted proteins. The cDNA sequences have been  
XX derived from a variety of human tissues. The invention also provides a  
XX method for producing proteins from these polynucleotide sequences. The  
XX proteins are useful for identifying compounds that modulate their  
XX activity and production, and the cell is also useful for identifying  
XX compounds that modulate expression of the polynucleotide sequences  
XX encoding the secreted proteins. The sequences of the invention are useful  
XX for treating diseases such as hyperproliferative disorders (e.g. cancer),  
XX immune deficiency disorders (e.g. severe combined immunodeficiency  
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
XX (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
XX infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
XX the invention are also useful in gene therapy. AAS62214-AAS62038  
XX represent the cDNA sequences of the invention that encode for novel human  
XX secreted proteins  
XX

SQ Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 6; Length 2176;  
Best Local Similarity 100.0%; Pred. No. 3e-233; Indels 0; Gaps 0;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGACGCGGTGAATGAGCGCGTGAAGAGCGAGTGCCTTCAATAGCTGAGTACC 60
DB 1205 AGGACGCGGTGAATGAGCGCGTGAAGAGCGAGTGCCTTCAATAGCTGAGTACC 1146
QY 61 CATGATGAAGAGCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120
DB 1145 CATGATGAAGAGCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 1086

```

QY 121 CACAGTGCATTATTAACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA 180  
DB 1085 CAAATGTCATTTATTAACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA 1026  
QY 181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 240  
DB 1025 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 966  
QY 241 CATCTAAGTGAACAAGAAACAGAAAGCCCAAGGCGATGCCAAGTGTCTATGAGA 300  
DB 965 CATCTAAGTGAACAAGAAACAGAAAGCCCAAGGCGATGCCAAGTGTCTATGAGA 906  
QY 301 CCCACCACTGCCAAGGCTCCGTGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 360  
DB 905 CCCACCACTGCCAAGGCTCCGTGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 846  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACGATATGCGGGGTGGTCTGCC 420  
DB 845 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACGATATGCGGGGTGGTCTGCC 786  
QY 421 ACCCGTGAAGGAGAGGAGCATGCGCACCACCTCCGTGAGGTCCAGAGAGCCAGAGAG 480  
DB 785 ACCCGTGAAGGAGAGGAGCATGCGCACCACCTCCGTGAGGTCCAGAGAGCCAGAGAG 726  
QY 481 TCTTGGGGGACCCATAGGTGTGCAATGAGAGGCGCTGGAAGATGAGAGAGCTTCCCTCC 540  
DB 725 TCTTGGGGGACCCATAGGTGTGCAATGAGAGGCGCTGGAAGATGAGAGAGCTTCCCTCC 666  
QY 541 AAGAGAGACCCCGGGGTTCCCGAAGGAAACCCCTCTGAGAGAGGAAAGTCCAGACCGAGC 600  
DB 665 AAGAGAGACCCCGGGGTTCCCGAAGGAAACCCCTCTGAGAGAGGAAAGTCCAGACCGAGC 606  
QY 601 TGGAGA CTGGCACTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGAGAAACAGA 660  
DB 605 TGGAGA CTGGCACTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGAGAAACAGA 546  
QY 661 GTGCAACCACTGTGAAGGCCCCAAAGCTTAAAGCTTCTCCCGCACCTTTCCGCCCC 720  
DB 545 GTGCAACCACTGTGAAGGCCCCAAAGCTTAAAGCTTCTCCCGCACCTTTCCGCCCC 486  
QY 721 GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGATGCGGGGAGAAAGAGTGGCCCTCAT 780  
DB 485 GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGATGCGGGGAGAAAGAGTGGCCCTCAT 426  
QY 781 GGAATGTGTGTGTCCCGGTGAAATGTTCAAGAGGTGGCCGTGTGTGAGACAGAGGTGCTT 840  
DB 425 GGAATGTGTGTGTCCCGGTGAAATGTTCAAGAGGTGGCCGTGTGTGAGACAGAGGTGCTT 366  
QY 841 CCGTGTGTGCGCGGGGCAATGAGCCGAGGTGGCTTGTGTGAGAAACAGAGGTGGCCCTTGG 900  
DB 365 CCGTGTGTGCGCGGGGCAATGAGCCGAGGTGGCTTGTGTGAGAAACAGAGGTGGCCCTTGG 306  
QY 901 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGTGTGAGAGACC 960  
DB 305 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGTGTGAGAGACC 246  
QY 961 TGGAAATATGATTAAGGCGAGACCGCTCAGAGCGCGAGAGATCGGCCCTTAC 1012  
DB 245 TGGAAATATGATTAAGGCGAGACCGCTCAGAGCGCGAGAGATCGGCCCTTAC 194

RESULT 3  
ADQ86032  
ID ADQ86032 standard; cDNA; 2372 BP.  
XX  
AC ADQ86032;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.  
XX  
KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;

KW cancer; cell proliferative disorder; gene; ss.  
XX Homo sapiens.  
OS  
PN MO2004060270-A2.  
XX  
XX 22-JUL-2004.  
PD  
XX  
PF 15-OCT-2003; 2003MO-US029126.  
XX  
PR 18-OCT-2002; 2002US-0418988P.  
XX  
XX (GENT ) GENENTECH INC.  
PA (WU/D/) WU T D.  
PA (ZHOU/) ZHOU Y.  
XX  
PI Wu TD, Zhou Y;  
XX  
DR WPI; 2004-534300/51.  
XX  
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,  
PT preventing or treating cell proliferative disorders such as cancer.  
XX  
PS Claim 1; SEQ ID NO 2904; 5504bp; English.

CC The present invention describes an isolated tumour-associated antigenic  
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of  
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
CC (c). Also described: (1) an expression vector comprising the above  
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
CC a process for producing a polypeptide; (4) an isolated polypeptide  
CC comprising: (a) an amino acid sequence encoded by any of the above  
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
CC length coding region of the above nucleotide sequences; or (c) a sequence  
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)  
CC an isolated antibody that binds to the above polypeptide; (7) a process  
CC for producing the antibody; (8) an isolated oligopeptide that binds to  
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
CC binding organic molecule that binds to the above polypeptide; (10) a  
CC composition of matter comprising the above (chimeric) polypeptide,  
CC antibody, oligopeptide or TAT binding organic molecule, in combination  
CC with a carrier; (11) an article of manufacture comprising a container and  
CC the composition of matter contained within the container; (12) methods of  
CC inhibiting the growth of a cell that expresses the above protein, where  
CC the growth of the cell is at least in part dependent upon a growth  
CC potentiating effect of the above protein; (13) a method of  
CC therapeutically treating a mammal having a cancerous tumour comprising  
CC cells that express the above protein; (14) a method of determining the  
CC presence of a protein in a sample suspected of containing the protein  
CC described above; (15) methods of diagnosing the presence of a tumour in a  
CC mammal; (16) a method for treating or preventing a cell proliferative  
CC disorder associated with increased expression or activity of the above  
CC protein; and (17) a method of binding an antibody, oligopeptide or  
CC organic molecule to a cell that expresses the protein described above.  
CC The TAT sequences have cytostatic activities, and can be used in gene  
CC therapy. The composition and methods are useful for diagnosing,  
CC preventing or treating cancer. The composition is also used for preparing  
CC a medicament for the therapeutic treatment or diagnostic detection of a  
CC cell proliferative disorder or cancer. The present sequence represents a  
CC human TAT cDNA sequence from the present invention.  
XX

SO Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 13; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 3, 1e-233;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGACCGGTGTAATGGGAGCGCTGTGAGAGCGAGGTGCTTCAATTAAGCTGTGGAGCC 60  
DB 1000 AGGACCGGTGTAATGGGAGCGCTGTGAGAGCGAGGTGCTTCAATTAAGCTGTGGAGCC 1059

QY 61 CATGATGAAAGACCAAGATCTTATCTAGAGCCCACTGTAGATCCAGATGAAAGACTCTGA 120  
 DB 1060 CATGATGAAAGACCAAGATCTTATCTAGAGCCCACTGTAGATCCAGATGAAAGACTCTGA 1119  
 QY 121 CAACAGTGAATTTATGTAACAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180  
 DB 1120 CAACAGTGAATTTATGTAACAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1179  
 QY 181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATCCA 240  
 DB 1180 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATCCA 1239  
 QY 241 CATCTACCTGGAACAAGAAACAAGAAAGCCAAAGGAGCCACAGTGTCTATGAGA 300  
 DB 1240 CATCTACCTGGAACAAGAAACAAGAAAGCCAAAGGAGCCACAGTGTCTATGAGA 1299  
 QY 301 CCCACCACTGCAAGAGCTCCGTGGAATGTTGATGAGAAAGATTTCAAGGAGCA 360  
 DB 1300 CCCACCACTGCAAGAGCTCCGTGGAATGTTGATGAGAAAGATTTCAAGGAGCA 1359  
 QY 361 ACTTAAAGTCTCTGCTGCGAAGAAAGCTTCAATGACAGTATGCGGGTGTCTGCC 420  
 DB 1360 ACTTAAAGTCTCTGCTGCGAAGAAAGCTTCAATGACAGTATGCGGGTGTCTGCC 1419  
 QY 421 ACCCGTGAAGGAGAGAGGATGCAACAACCTCCGTGAGTCCAGAGGAGGAGG 480  
 DB 1420 ACCCGTGAAGGAGAGAGGATGCAACAACCTCCGTGAGTCCAGAGGAGGAGG 1479  
 QY 481 TCCGTGGGAGACCATGAGGTGCGATGAGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1480 TCCGTGGGAGACCATGAGGTGCGATGAGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 1539  
 QY 541 AAGAAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCAAGACCGAGC 600  
 DB 1540 AAGAAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCAAGACCGAGC 1599  
 QY 601 TGAAGACTGAGCAGTGTCCCAATCCGGGTGTGAGAAACAACACTTGGCCGAGAAACAGA 660  
 DB 1600 TGAAGACTGAGCAGTGTCCCAATCCGGGTGTGAGAAACAACACTTGGCCGAGAAACAGA 1659  
 QY 661 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAAGGCTTCTCCGCCACCTTTCCGCCCCC 720  
 DB 1660 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAAGGCTTCTCCGCCACCTTTCCGCCCCC 1719  
 QY 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATGCGGGAGAGAAAGGTGCTCAT 780  
 DB 1720 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATGCGGGAGAGAAAGGTGCTCAT 1779  
 QY 781 GGAATCGTGTGTCCCGGTGGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 840  
 DB 1780 GGAATCGTGTGTCCCGGTGGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 1839  
 QY 841 CCGTGTGTGCGGGGAGTGAACCGAGGTGCTTGTGTGAGAGAAAGAGTGGCCCTG 900  
 DB 1840 CCGTGTGTGCGGGGAGTGAACCGAGGTGCTTGTGTGAGAGAAAGAGTGGCCCTG 1899  
 QY 901 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGAGAGACC 960  
 DB 1900 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGCTGAGAGACC 1959  
 QY 961 TGAAGAAATGATGAAG 1012  
 DB 1960 TGAAGAAATGATGAAG 2011

## RESULT 4

ID AD516277 standard; cDNA; 2390 BP.  
 AC AD516277;  
 XX  
 DT 18-NOV-2004 (first entry)

XX DE Human cDNA encoding the Ewing sarcoma protein SegID 1.  
 XX KW human; sg; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
 XX KW hypogonadism; androgen-resistance syndrome; testicular feminisation.  
 XX OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 44..2014  
 FT /tag= a  
 FT /product= "Ewing sarcoma protein"  
 XX  
 PN EPI455190-A1.  
 XX  
 PD 08-SEP-2004.  
 XX  
 PF 16-FEB-2004; 2004EP-00003422.  
 XX  
 PR 04-MAR-2003; 2003DE-01009280.  
 PR 25-APR-2003; 2003US-0465692P.  
 XX  
 PA (SCHD); SCHERING AG.  
 PI Obendorf M, Wolf S;  
 XX  
 DR WPI; 2004-627861/61.  
 DR P-PSDB; AD516278.  
 XX  
 PT Determining the hormonal effects of substances, used to identify  
 PT pharmaceutical, e.g. for treatment of androgen receptor dysfunction,  
 PT from modulating interaction between nuclear receptors and Ewing sarcoma  
 PT protein.  
 XX  
 PS Claim 7; SEQ ID NO 1; 30pp; German.  
 XX  
 CC This invention relates to a novel modulators that alter the interaction  
 CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
 CC as the screening method thereof. Specifically, it refers to determining  
 CC and identifying a hormonal effect brought about by test compounds that  
 CC modulate either the binding of EWS to the nuclear receptor or the ligand-  
 CC induced activity of this receptor. The present invention describes the  
 CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
 CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
 CC receptors. Accordingly, these modulators may be used in the development  
 CC of pharmaceutical compositions that can diagnose and be used to treat  
 CC diseases associated with receptor dysfunction such as prostatic cancer,  
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
 CC testicular feminisation. This method provides reliable, sensitive,  
 CC simple, inexpensive and rapid assessment of the hormonal effects of these  
 CC test compounds. This polynucleotide sequence is the cDNA encoding the  
 CC human Ewing sarcoma protein of the invention.  
 XX  
 SQ Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 1012; DB 13; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-233;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGACCCGGTGAATGAGAGAGCGCTGAGAGAGAGGTGCTTCAATGAAGCTGTGAGACC 60  
 DB 1000 AGGACCCGGTGAATGAGAGAGCGCTGAGAGAGAGGTGCTTCAATGAAGCTGTGAGACC 1059  
 QY 61 CATGATGAAAGACCAAGATCTTATCTAGAGCCCACTGTAGATCCAGATGAAAGACTCTGA 120  
 DB 1060 CATGATGAAAGACCAAGATCTTATCTAGAGCCCACTGTAGATCCAGATGAAAGACTCTGA 1119  
 QY 121 CAACAGTGAATTTATGTAACAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180  
 DB 1120 CAACAGTGAATTTATGTAACAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1179  
 QY CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATCCA 240

Db 1180 CTTCTTTAAGCAGTGTGGGGTTTGAATGATGAACAGAGAACTGGGCAACCATGATCCA 1239  
 QY 241 CATCTACTGTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 300  
 Db 1240 CATCTACTGTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 1299  
 QY 301 CCCACCACTGCGCAAGGCTGCGTGAATGATGAGGAAAGATTTTCAAGGAGACAA 360  
 Db 1300 CCCACCACTGCGCAAGGCTGCGTGAATGATGAGGAAAGATTTTCAAGGAGACAA 1359  
 QY 361 ACTTAAAGTCTCCCTTCTCGAAGAAAGCTCCAGTGAAGATATGCGGGGTGCTCTGCC 420  
 Db 1360 ACTTAAAGTCTCCCTTCTCGAAGAAAGCTCCAGTGAAGATATGCGGGGTGCTCTGCC 1419  
 QY 421 ACCCGGTGAGGCGAGAGGATGCCACCACTTCCTGTGAGGTCCAGAGGCCCAAGAGG 480  
 Db 1420 ACCCGGTGAGGCGAGAGGATGCCACCACTTCCTGTGAGGTCCAGAGGCCCAAGAGG 1479  
 QY 481 TCCGTGGGGGAGCCATGGGGTGGCATGGAGGCCGCTGGAGAGATGAGAGAGGCTTCCCTCC 540  
 Db 1480 TCCGTGGGGGAGCCATGGGGTGGCATGGAGGCCGCTGGAGAGATGAGAGAGGCTTCCCTCC 1539  
 QY 541 AAGAGACCCCGGGGGTTCGCCAGAGGAAACCCCTCTGAGAGAGAAACGTCACGACGAGC 600  
 Db 1540 AAGAGACCCCGGGGGTTCGCCAGAGGAAACCCCTCTGAGAGAGAAACGTCACGACGAGC 1599  
 QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACAGAACTTCGCTGAGAAACAGA 660  
 Db 1600 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACAGAACTTCGCTGAGAAACAGA 1659  
 QY 661 GTGCAACCACTGTAAAGGCCCAAAAGCCTGAAGGCTTCTCCGCAACCTTCCGCCCC 720  
 Db 1660 GTGCAACCACTGTAAAGGCCCAAAAGCCTGAAGGCTTCTCCGCAACCTTCCGCCCC 1719  
 QY 721 GGGTGTGTGATCTGTGACAGAGTGGCCCTGTGTGCATCGGGAGAGAAAGTGGCTCAT 780  
 Db 1720 GGGTGTGTGATCTGTGACAGAGTGGCCCTGTGTGCATCGGGAGAGAAAGTGGCTCAT 1779  
 QY 781 GGATCTGTGTGTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGGTGGCTT 840  
 Db 1780 GGATCTGTGTGTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGGTGGCTT 1839  
 QY 841 CCGTGTGTGGCCGGGCGATGAGACCGAGGTGGCTTGTGTGAGAGAAAGACAGTGGCTTGG 900  
 Db 1840 CCGTGTGTGGCCGGGCGATGAGACCGAGGTGGCTTGTGTGAGAGAAAGACAGTGGCTTGG 1899  
 QY 901 GGGGCCCCCTTGACCTTTGATGGAACAGATGGGAGAGAAAGAGAGAGAGTGGAGGACC 960  
 Db 1900 GGGGCCCCCTTGACCTTTGATGGAACAGATGGGAGAGAAAGAGAGAGAGTGGAGGACC 1959  
 QY 961 TGGAAAAATGGAATAAAGGCGAGACCGGTCAAGAGCGGAGAGATCGGCTTAC 1012  
 Db 1960 TGGAAAAATGGAATAAAGGCGAGACCGGTCAAGAGCGGAGAGATCGGCTTAC 2011  
 RESULT 5  
 AAQ50643  
 ID AAQ50643 standard; cDNA; 2371 BP.  
 XX  
 AC AAQ50643;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 26-MAY-1994 (first entry)  
 XX  
 DE Human Ews gene clone Bf1AC5 from foetal brain cDNA library.  
 XX  
 KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;  
 KW malignant melanoma; hum-fli-1;  
 KW primitive peripheral neuroectodermal tumour; human chromosome 11;  
 KW human chromosome 22; ss.  
 XX  
 OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT CDS 25..1995  
 FT /tag= a  
 FT /product= "EWS protein"  
 FT /transl\_except= pos:1729..1731; aa:Val  
 FT polyA\_signal  
 FT /tag= b  
 FT polyA\_signal 2331..2336  
 FT /tag= c  
 XX  
 XX MO9323549-A2.  
 XX  
 XX 25-NOV-1993.  
 XX  
 XX 19-MAY-1993; 93MO-FR000494.  
 XX  
 XX 20-MAY-1992; 92FR-00006123.  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougastel B,  
 PI Thomas G, Zucman J;  
 XX  
 XX WPI: 1993-386580/48.  
 XX  
 XX P-PSDB; AAR44555.  
 XX  
 XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion  
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
 XX  
 XX Disclosure; Fig 6; 123p; French.  
 XX  
 XX The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA  
 CC library (Stratagene cat. # 936206). The clone Bf1AC5 was identified and  
 CC sequenced. It represents the entire coding region and 3'-UTR of the Ews  
 CC gene. (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 1010.4; DB 2; Length 2371;  
 Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGGAGCGGTGGAATGAGGCGGCTGAGAGCGAGGTGGCTTCAATAGCTGTGGACC 60  
 Db 981 AGGAGCGGTGGAATGAGGCGGCTGAGAGCGAGGTGGCTTCAATAGCTGTGGACC 1040  
 QY 61 CATGATGAAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGAATCTGA 120  
 Db 1041 CATGATGAAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGAATCTGA 1100  
 QY 121 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTGACTCTGATGATGTGGCAGA 180  
 Db 1101 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTGACTCTGATGATGTGGCAGA 1160  
 QY 181 CTTCTTTAAGCAGTGTGGGGTTTGAATGAAACAAGAACTGGCAACCCATGATCCA 240  
 Db 1161 CTTCTTTAAGCAGTGTGGGGTTTGAATGAAACAAGAACTGGCAACCCATGATCCA 1220  
 QY 241 CATCTACTGTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 300  
 Db 1221 CATCTACTGTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 1280  
 QY 301 CCCACCACTGCGCAAGGCTGCGTGAATGATGAGGAAAGATTTTCAAGGAGACAA 360  
 Db 1281 CCCACCACTGCGCAAGGCTGCGTGAATGATGAGGAAAGATTTTCAAGGAGACAA 1340  
 QY 361 ACTTAAAGTCTCCCTTCTCGAAGAAAGCTCCAGTGAAGATATGCGGGGTGCTCTGCC 420  
 Db 1341 ACTTAAAGTCTCCCTTCTCGAAGAAAGCTCCAGTGAAGATATGCGGGGTGCTCTGCC 1400  
 QY 421 ACCCGGTGAGGCGAGAGGATGCCACCACTTCCTGTGAGGTCCAGAGGCCCAAGAGG 480

Db	1401	ACCCTGAGAGGACGAGGCAATGCCACCACTCCTCGTAGAGTGTCCAGAGGACCCAGAGG	1460
Qy	481	TCCTGGGGGAGACCCATGGGTCGCATGGGAGGCGCTGTGAGAGAGATGAGAGGGCTTCCTCC	540
Db	1461	TCCCTGGGGGAGACCCATGGGTCGCATGGGAGGCGCTGTGAGAGAGATGAGAGGGCTTCCTCC	1520
Qy	541	AAGAGGAGCCCCGGGGGTTCCCGAGGGGAAACCCCTCTGAGAGAGGAAACGTCAGCACCGAGC	600
Db	1521	AAGAGGAGCCCCGGGGGTTCCCGAGGGGAAACCCCTCTGAGAGAGGAAACGTCAGCACCGAGC	1580
Qy	601	TGAGAGCTGGCAGTGTCCTCCCAATCCCGGTTGTGAAACCAAGAACTTCGCTTGAGAAACAGA	660
Db	1581	TGAGAGCTGGCAGTGTCCTCCCAATCCCGGTTGTGAAACCAAGAACTTCGCTTGAGAAACAGA	1640
Qy	661	GTGCACCAACGTGTAAAGGCCCCCAAGAGCTGAAAGGCTTCTCCCGCACCCTTCCGGCCCC	720
Db	1641	GTGCACCAACGTGTAAAGGCCCCCAAGAGCTGAAAGGCTTCTCCCGCACCCTTTCGGCCCC	1700
Qy	721	GGGTGTGTGATCGTGGCAGAGGTGGCCCTGTGTGGCATGTGGGGAGAGAGAGGTGGCTCAT	780
Db	1701	GGGTGTGTGATCGTGGCAGAGGTGGCCCTGTGTGGCATGTGGGGAGAGAGAGGTGGCTCAT	1760
Qy	781	GGATCGTGTGATGTCCTCCGCTGGAAATGTTCAAGAGTGGCCCTGTGTGAGACACAGAGTGGCTT	840
Db	1761	GGATCGTGTGATGTCCTCCGCTGGAAATGTTCAAGAGTGGCCCTGTGTGAGACACAGAGTGGCTT	1820
Qy	841	CCGTGTGTGGCCGGGGGACATGAAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCTGG	900
Db	1821	CCGTGTGTGGCCGGGGGACATGAAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCTGG	1880
Qy	901	GGGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGGAAGAGAGAGAGAGAGAGAGAGAGC	960
Db	1881	GGGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGGAAGAGAGAGAGAGAGAGAGAGAGC	1940
Qy	961	TGGAAGAAATGAGTAAAGCGCAGCAACCGTCAGAGCGCAGAGATCGCCCTTAC	1012
Db	1941	TGGAAGAAATGAGTAAAGCGCAGCAACCGTCAGAGCGCAGAGATCGCCCTTAC	1992
RESULT 6			
AAS70647			
ID AAS70647 standard; cDNA; 2388 BP.			
AC AAS70647;			
DT 13-FEB-2002 (first entry)			
DE DNA encoding novel human diagnostic protein #6451.			
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;			
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.			
OS Homo sapiens.			
PN WO200175067-A2.			
PD 11-OCT-2001.			
PF 30-MAR-2001; 2001WO-US008631.			
PR 31-MAR-2000; 2000US-00540217.			
PR 23-AUG-2000; 2000US-00649167.			
PA (HYSE-) HYSEQ INC.			
PI Drmanac RT, Liu C, Tang YT;			
DR WPI; 2001-639362/73.			
XX P-PSDB; ABG06460.			
PT New isolated polynucleotide and encoded polypeptides, useful in			
PT diagnostics, forensics, gene mapping, identification of mutations			
PT responsible for genetic disorders or other traits and to assess			

PT	biodiversity.
XX	
PS	Claim 1; SEQ ID NO 6451; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC	sequences. (I) is useful as hybridisation probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (I) is useful in gene therapy techniques to restore normal
CC	activity of (II) or to treat disease states involving (II). (II) is
CC	useful for generating antibodies against it, detecting or quantifying a
CC	polypeptide in tissue, as molecular weight markers and as a food
CC	supplement. (II) and its binding partners are useful in medical imaging
CC	of sites expressing (II). (I) and (II) are useful for treating disorders
CC	involving aberrant protein expression or biological activity. The
CC	polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC	coding sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 2388 BP; 628 A; 599 C; 668 G; 493 T; 0 U; 0 Other;
Query Match	99.8%; Score 1010.4; DB 5; Length 2388;
Best Local Similarity	99.9%; Pred. No. 7.5e-233;
Matches 1011; Conservative %	0; Mismatches 1; Indels 0; Gaps 0;
OY	1 AGAGCGGGGTGGAATGGGACCGCTGGAGAGCAGGTGCCTCAATAAGCTGGTGGACC 60
DB	1017 AGAGCGGGGTGGAATGGGACCGCTGGAGAGCGGTGCTTCATAAAGCTGGTGGACC 1076
OY	61 CATGATGAAGAAGCACCAATCTTGATCTAAGGCCCACTGTAGATCCAGATGAAGACTGTA 120
DB	1077 CATGATGAAGAAGCACCAATCTTGATCTAAGGCCCTCTGTAGATCCAGATGAAGACTGTA 1136
OY	121 CAACAGTGCAATTATGTACAAGGATTAAATGACAGTGTGACTGTAGATGATCTGGCAGA 180
DB	1137 CAACAGTGCAATTATGTACAAGGATTAAATGACAGTGTGACTGTAGATGATCTGGCAGA 1196
OY	181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 240
DB	1197 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 1256
OY	241 CATCTACCTGGACAAAGAAACAGAAAAGCCCAAAGGGGATGCCACAGTGTCTTATGAAGA 300
DB	1257 CATCTACCTGGACAAAGAAACAGAAAAGCCCAAAGGGGATGCCACAGTGTCTTATGAAGA 1316
OY	301 CCACACCATCTGCCAAGGCTGCCGTGGAATGGTTTGTATGGGAAGATTTCAAGGAGACAA 360
DB	1317 CCACACCATCTGCCAAGGCTGCCGTGGAATGGTTTGTATGGGAAGATTTTCAAGGAGACAA 1376
OY	361 ACTTAAAGTCTCCCTTGCTGGGAAGAAGCTTCAATGAACAGTATGGGGGTGTCTGCC 420
DB	1377 ACTTAAAGTCTCCCTTGCTGGGAAGAAGCTTCAATGAACAGTATGGGGGTGTCTGCC 1436
OY	421 ACCCGTGAAGGGCAGAGGCATGCCAACCACTCCGTGAGAGGTCCAGAGGCCCAAGAGG 480
DB	1437 ACCCGTGAAGGGCAGAGGCATGCCAACCACTCCGTGAGAGGTCCAGAGGCCCAAGAGG 1496
OY	481 TCCTGGGGGGAACCAATGGGTGCGATGGGAGGCGCTGTGAGAGAGATAGAGAGGCTTCCCTCC 540
DB	1487 TCCTGGGGGGAACCAATGGGTGCGATGGGAGGCGCTGTGAGAGAGATAGAGAGGCTTCCCTCC 1556
OY	541 AAAGAGACCCCGGGGTTTCCGAGGGGAACCCCTTGTGAGAGAGAAAGTCCACAGACCGAGC 600
DB	1557 AAAGAGACCCCGGGGTTTCCGAGGGGAACCCCTTGTGAGAGAGAAAGTCCACAGACCGAGC 1614
OY	601 TGAGACTGCGACAGTCTCCCAATCCGGGTTGTGGAAACCAAGAACTTCGCTGGAAGACAGA 660

Db 1617 TGAAGACTGACAGTGTCCCAATCCGGTGTGTGAAACCAAACTTCCCTGAGAAACGA 1676  
 Qy 661 GTGCACACCACTGTAGAGCCCAAAAGCTTGAAGGCTTCTCCCGCACCTTTCCGCCCC 720  
 Db 1677 GTCCAAACCACTGTAGAGCCCAAAAGCTTGAAGGCTTCTCCCGCACCTTTCCGCCCC 1736  
 Qy 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCAATCCGGGAGAGAAAGAGTGGCTCAT 780  
 Db 1737 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCAATCCGGGAGAGAAAGAGTGGCTCAT 1796  
 Qy 781 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGTGAGAGAGAGTGGCTT 840  
 Db 1797 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGTGAGAGAGAGTGGCTT 1856  
 Qy 841 CCGTGTGTGTCGGGGGAGTGAACCGAGGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGG 900  
 Db 1857 CCGTGTGTGTCGGGGGAGTGAACCGAGGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGG 1916  
 Qy 901 GGGGCCCCCTTGACCTTTGATGAAACAGATGTGGAGAGAGAGAGAGAGTGGAGACC 960  
 Db 1917 GGGGCCCCCTTGACCTTTGATGAAACAGATGTGGAGAGAGAGAGAGAGTGGAGACC 1976  
 Qy 961 TGGAAAAATGATTAAGGAGAGACCGTCAAGAGCGGAGAGATCGGCTTAC 1012  
 Db 1977 TGGAAAAATGATTAAGGAGAGACCGTCAAGAGCGGAGAGATCGGCTTAC 2028  
 RESULT 7  
 ID ABR84628 standard; cDNA; 2390 BP.  
 AC ABR84628;  
 XX  
 DT 14-AUG-2002 (first entry)  
 DE Human cDNA differentially expressed in granulocytic cells #1199.  
 XX  
 KW Human; B6; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; peridontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001; 2001MO-US030821.  
 XX  
 PR 03-OCT-2000; 2000US-0237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;  
 XX  
 DR WPI; 2002-43528/46.  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.  
 XX  
 PS Claim 1; SEQ ID NO 1199; 114bp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where

CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) Gs by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease using the gene expression  
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease, by detecting the level of  
 CC expression in a sample of the tissue of gene(s) from Gs, where the level  
 CC of expression of the gene is indicative of inflammation; (4) treating  
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease, by contacting a tissue having inflammation with an  
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
 CC is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful  
 CC for screening an agent capable of modulating GCA preferably in an  
 CC inflammation in a tissue; M4 is useful for detecting an inflammation  
 CC (especially chronic) in a tissue, an allergic response in a subject,  
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
 CC disease, ulcerative colitis, peridontal disease, also bacterial  
 CC infection, viral infection, parasitic infection, protozoal infection,  
 CC fungal infection and M5 is useful for treating one of the above  
 CC conditions. The present sequence represents a gene differentially  
 CC expressed in granulocytes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 1010.4; DB 6; Length 2390;  
 Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 AGGAGCGGCTGAGTATGAGGAGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 60  
 Db 1000 AGGAGCGGCTGAGTATGAGGAGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGTGGACC 1059  
 Qy 61 CATGATGAAAGACCAATCTTATCTAGGCCCCCACTGTATGATCCAGATTAAGACTTGA 120  
 Db 1060 CATGATGAAAGACCAATCTTATCTAGGCCCCCACTGTATGATCCAGATTAAGACTTGA 1119  
 Qy 121 CAACAGTGCATTTATGTACAAGATTAATGACAGTGTACTCTGTATGATCTGGCAGA 180  
 Db 1120 CAACAGTGCATTTATGTACAAGATTAATGACAGTGTACTCTGTATGATCTGGCAGA 1179  
 Qy 181 CTTCTTTAAGCACTGTGTGGGTTGTAAATGAAACAAGAACTGGCAACCCATGATCCA 240  
 Db 1180 CTTCTTTAAGCACTGTGTGGGTTGTAAATGAAACAAGAACTGGCAACCCATGATCCA 1239  
 Qy 241 CATCTACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA 300  
 Db 1240 CATCTACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA 1299  
 Qy 301 CCCACCACATGCGCAAGGCTGCGGTGAATGTGTGAATGGGAAAGATTTTCAAGGAGCAA 360  
 Db 1300 CCCACCACATGCGCAAGGCTGCGGTGAATGTGTGAATGGGAAAGATTTTCAAGGAGCAA 1359  
 Qy 361 ACTTAAGTCTCCCTTGTGGAAGAGCTCCCAATGAACATGATCGGGGTGTGCC 420  
 Db 1360 ACTTAAGTCTCCCTTGTGGAAGAGCTCCCAATGAACATGATCGGGGTGTGCC 1419  
 Qy 421 ACCCGCTGAGGAGGAGGATGCAACCACTCCGTGTGAGAGTCCAGAGGCCCAAGAG 480  
 Db 1420 ACCCGCTGAGGAGGAGGATGCAACCACTCCGTGTGAGAGTCCAGAGGCCCAAGAG 1479  
 Qy 481 TCTGTGGGAGACCATGAGTGTGAGAGAGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 540  
 Db 1480 TCTGTGGGAGACCATGAGTGTGAGAGAGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 1539

OY	541	AAGAGAACCCGGGGATTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCACGACCCGAGC	600
Db	1540	AAGAGAACCCGGGGATTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCACGACCCGAGC	1539
OY	601	TGAGACTGCGAGTGTGCCAATCCGGGGTGTGAGAAACGAACTTCGCTTGAGAGACGA	660
Db	1600	TGAGAGACTGCGAGTGTGCCAATCCGGGGTGTGAGAAACGAACTTCGCTTGAGAGACGA	1659
OY	661	GTGACACCAAGTGTAAAGCCCCAAAGCCGTAAAGGCTTCTCCCGCACCCCTTCCGGCCCC	720
Db	1660	GTGACACCAAGTGTAAAGCCCCAAAGCCGTAAAGGCTTCTCCCGCACCCCTTCCGGCCCC	1719
OY	721	GCGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATGCGGGGAGAGAAAGGTGGCTTCAT	780
Db	1720	GCGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATGCGGGGAGAGAAAGGTGGCTTCAT	1779
OY	781	GGATCTGTGTGTGGTCCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAGACAAGATGTGGTT	840
Db	1780	GGATCTGTGTGTGGTCCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAGACAAGATGTGGTT	1839
OY	841	CCGTGTGTGGCGGGGCATGGAACCGAGTGTGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	900
Db	1840	CCGTGTGTGGCGGGGCATGGAACCGAGTGTGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	1839
OY	901	GGGGCCCCCTTGACCTTTGATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGACCC	960
Db	1900	GGGGCCCCCTTGACCTTTGATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGACCC	1959
OY	961	TGGAATAATGTGTTAAAGCGAGCAACCGTCAGGAGCGCAGAGTGGGCCCTAC	1012
Db	1960	TGGAATAATGTGTTAAAGCGAGCAACCGTCAGGAGCGCAGAGTGGGCCCTAC	2011

## RESULT 8

ID	ABK64822 standard; DNA; 2390 BP.
XX	
AC	ABK64822;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Human benign prostatic hyperplasia gene #717.
XX	
KM	Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200212440-A2.
XX	
PD	14-FEB-2002.
XX	
PF	07-AUG-2001; 2001WO-US024708.
XX	
PR	07-AUG-2000; 2000US-0223323P.
XX	05-JUN-2001; 2001US-00873319.
XX	
XX	(GENE-) GENE LOGIC INC.
PA	(NISB ) JAPAN TOBACCO INC.
XX	
PI	Munger WE, Kulikarni P, Getzenberg RH, Waga I, Yamamoto J;
DR	WPI; 2002-257476/30.
XX	
PT	Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT	detecting expression levels of one or more genes in prostate cells from
PT	patient that are differentially regulated compared to normal prostate
PT	cells.
XX	
PS	Disclosure; Page 399-400; 444pp; English.
XX	
CC	The invention relates to a method of diagnosing (I) the onset or
CC	progression of benign prostatic hyperplasia (BPH), or screening (II) for

Query Match	99.8%	Score 1010.4	DB 6	Length 2390
Best Local Similarity	99.9%	Pred. No. 7.5e-233		
Matches 1011	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	AGGACGCGGTGAAATGAGGACGCGTGGAGAGCGAGGTGGCTTCATTAAGCTTGTTGAC	60	
Db	1000	AGGACGCGGTGAAATGAGGACGCGTGGAGAGCGAGGTGGCTTCATTAAGCTTGTTGAC	1059	
QY	61	CATGAGTAAAGACACAAATCTTATCTTAAGCCCACTGTAGATCAATGAAGCTCTGA	120	
Db	1060	CATGAGTAAAGACACAAATCTTATCTTAAGCCCTCTGTAGATCAATGAAGCTCTGA	1119	
QY	121	CAACAGTGCATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATCTGTGCAGA	180	
Db	1120	CAACAGTGCATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATCTGTGCAGA	1179	
QY	181	CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA	240	
Db	1180	CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA	1239	
QY	241	CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCACAGTGTCTTATGAGA	300	
Db	1240	CATCTACTGGAACAAGAAACAGAAAGCCCAAGGGGATGCCACAGTGTCTTATGAGA	1299	
QY	301	CCCAACCACTCCCAAGGCTGCGTGAATGTTTGAATGGGAAAGTTTCAAGGAGCA	360	
Db	1300	CCCAACCACTCCCAAGGCTGCGTGAATGTTTGAATGGGAAAGTTTCAAGGAGCA	1359	
QY	361	ACTTAAAGTCTCCCTGTGCGAAGAAAGCTTCAATGAACGTAATGCGGGGTGTCTGCC	420	
Db	1360	ACTTAAAGTCTCCCTGTGCGAAGAAAGCTTCAATGAACGTAATGCGGGGTGTCTGCC	1419	
QY	421	ACCCGTGAGGGCAAGGCAATGCCAACCACTCCGTGAGAGTCCAGAGGCCAGAGAG	480	
Db	1420	ACCCGTGAGGGCAAGGCAATGCCAACCACTCCGTGAGAGTCCAGAGGCCAGAGAG	1479	
QY	481	TCTGGGGGGAACCAATGGGTGCGATGGGAGGCGGTGAGAGAAATGAGAGGCTTCCCTCC	540	
Db	1480	TCTGGGGGGAACCAATGGGTGCGATGGGAGGCGGTGAGAGAAATGAGAGGCTTCCCTCC	1539	
QY	541	AAGAGAACCCGGGGATTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC	600	
Db	1540	AAGAGAACCCGGGGATTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC	1599	
QY	601	TGAGAGCTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAACTTCCGCTGAGAAACGA	660	
Db	1600	TGAGAGCTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAACTTCCGCTGAGAAACGA	1659	
QY	661	GTGCAACCAAGTGAAGGCCCAAGGCTGAAGGCTTCTCCGGCACCCCTTCCGGCCCC	720	



Db 1660 GTGCACCAAGTGAAGGCCCCAAGCCTGAAGGCTTCCTCCGCCACCCCTTTCCGCCCCC 1719  
Qy 721 GGGTGTGATCTGTGGCAGAGTGGCCCTGTGTGCAATGCGGGAGAGAGAGTGGCTTCAT 780  
Db 1720 GGGGTGTGATCTGTGGCAGAGTGGCCCTGTGTGCAATGCGGGAGAGAGAGTGGCTTCAT 1779  
Qy 781 GGATCGTGTGTCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 840  
Db 1780 GGATCGTGTGTCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 1839  
Qy 841 CCGTGTGTGCGCGGGGCAATGACCCGAGGTGTGCTTTGTGTGAGAGAGAGAGTGGCCCTG 900  
Db 1840 CCGTGTGTGCGCGGGGCAATGACCCGAGGTGTGCTTTGTGTGAGAGAGAGAGTGGCCCTG 1899  
Qy 901 GGGGCCCCCTGTGACCTTTGATGAACTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 1900 GGGGCCCCCTGTGACCTTTGATGAACTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959  
Qy 961 TGGAAAAATGATTAAGGCGAGCAGCGTCAAGAGAGAGAGAGATGCGCCCTAC 1012  
Db 1960 TGGAAAAATGATTAAGGCGAGCAGCGTCAAGAGAGAGAGAGATGCGCCCTAC 2011

## RESULT 9

ABN97274 ID ABN97274 standard, DNA, 2390 BP.

ABN97274;

13-AUG-2002 (first entry)

Gene #3772 used to diagnose liver cancer.

Gene; liver cancer; de; hepatocellular carcinoma; hepatocytic;  
metastatic liver tumor; cytostatic; expression profile; disease state;  
disease progression; drug toxicity; drug efficacy; drug metabolism.

Homo sapiens.

MO200229103-A2.

11-APR-2002.

02-OCT-2001; 2001MO-US030589.

02-OCT-2000; 2000US-0237054P.

(GENE-) GENE LOGIC INC.

Horne D, Alvares C, Perez-Da-Silva S, Vockley JG;

WPI; 2002-426119/45.

Diagnosing and detecting the progression of liver cancer, hepatocellular  
carcinoma or metastatic liver tumor in a patient, involves detecting the  
level of expression of two or more genes in a liver tissue sample.

Claim 1; SEQ ID NO 3772; 298bp; English.

The invention relates to a novel method for diagnosing and detecting the  
progression of liver cancer, hepatocellular carcinoma or metastatic liver  
tumor in a patient, and differentiating metastatic liver cancer from  
hepatocellular carcinoma in a patient, involving detecting the level of  
expression of two or more genes represented in ABN93503-ABN97455 in a  
tissue sample. The method of the invention has hepatocytic, and  
cytostatic activity. The method is useful for diagnosing and detecting  
the progression of liver cancer, hepatocellular carcinoma and metastatic  
liver carcinoma in a patient. The method is useful for identifying  
expression profiles which serve as useful diagnostic markers as well as  
markers that can be used to monitor disease states, disease progression,  
drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
for this patent did not form part of the printed specification, but was  
obtained in electronic format directly from WIPO at

ftp.wipo.int/pub/published\_pct\_sequences

Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Query Match 99.8%; Score 1010.4; DB 6; Length 2390;

Best Local Similarity 99.9%; Pred. No. 7.5e-233;

Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAGCGGCTGTGAATGGGCAAGCCCTGTGAGAGCGAGGTGGCTTCAATAGACCTGTGTGACC 60  
Db 1000 AGAGCGGCTGTGAATGGGCAAGCCCTGTGAGAGCGAGGTGGCTTCAATAGACCTGTGTGACC 1059  
Qy 61 CATGAG 120  
Db 1060 CATGAG 1119  
Qy 121 CAACAGTGCATTTATTAATGAAGATTAATGAAGATTAATGAAGATTAATGAAGATTAATGAAG 180  
Db 1120 CAACAGTGCATTTATTAATGAAGATTAATGAAGATTAATGAAGATTAATGAAGATTAATGA 1179  
Qy 181 CTTCTTTAAGCAAGTGTGGGCTTTGTTAAGATGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
Db 1180 CTTCTTTAAGCAAGTGTGGGCTTTGTTAAGATGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239  
Qy 241 CATCTACCTGTGACAG 300  
Db 1240 CATCTACCTGTGACAG 1239  
Qy 301 CCCACCACTGCGCAAGGCTGCGGTGAAATGTTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Db 1300 CCCACCACTGCGCAAGGCTGCGGTGAAATGTTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359  
Qy 361 ACTTAAAGTCTCCCTTGTCTGTGAG 420  
Db 1360 ACTTAAAGTCTCCCTTGTCTGTGAG 1419  
Qy 421 ACCCGGTGAG 480  
Db 1420 ACCCGGTGAG 1479  
Qy 481 TCCGTGGGAG 540  
Db 1480 TCCGTGGGAG 1539  
Qy 541 AAG 600  
Db 1540 AAG 1599  
Qy 601 TGGAGACTGGCAATGTTCCCAATCCGGGTGTGAAACAGAACTTCCCTGAGAGACAGA 660  
Db 1600 TGGAGACTGGCAATGTTCCCAATCCGGGTGTGAAACAGAACTTCCCTGAGAGACAGA 1659  
Qy 661 GTGCACCAAGTGAAGGCCCCAAGGCTGAAGGCTTCTCCGCAACCTTCCGCCCC 720  
Db 1660 GTGCACCAAGTGAAGGCCCCAAGGCTGAAGGCTTCTCCGCAACCTTCCGCCCC 1719  
Qy 721 GGGTGTGATCTGTGGCAGAGTGGCCCTGTGTGCAATGCGGGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 1720 GGGTGTGATCTGTGGCAGAGTGGCCCTGTGTGCAATGCGGGAGAGAGAGAGAGAGAGAGAGAG 1779  
Qy 781 GGATCGTGTGTCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 840  
Db 1780 GGATCGTGTGTCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 1839  
Qy 841 CCGTGTGTGCGCGGGGCAATGACCCGAGGTGTGCTTTGTGTGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 1840 CCGTGTGTGCGCGGGGCAATGACCCGAGGTGTGCTTTGTGTGAGAGAGAGAGAGAGAGAGAGAG 1899  
Qy 901 GGGGCCCCCTGTGACCTTTGATGAACTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 1900 GGGGCCCCCTGTGACCTTTGATGAACTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959  
Qy 961 TGGAAAAATGATTAAGGCGAGCAGCGTCAAGAGAGAGAGAGATGCGCCCTAC 1012

DB 1660 TGAATAATGATTAAGGCGAGCACTCTCAGAGCCGAGATCGCCCTTAC 2011

RESULT 10  
ACN40903  
ID ACN40903 standard; cDNA; 2390 BP.

AC ACN40903;

DT 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.

XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX chromosome identification; chromosome mapping; gene mapping;  
XX gene therapy; cytostatic; gene; 88.

OS Homo sapiens.

XX WO2004030615-A2.

PN 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-0414971P.

XX (GENTH) GENTECH INC.

XX Wu TD, Zhang Z, Zhou Y;

XX WPI; 2004-347921/32.

XX P-PSDB; ABM62330.

PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.

XX Claim 1; SEQ ID NO 5989; 7273bp; English.

XX The invention relates to human tumor-associated antigenic target (TAT)  
XX polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX overexpressed in cancer tissues compared to normal tissues, and may thus  
XX serve as effective targets for the diagnosis and treatment of cancer in  
XX mammals. The invention also relates to nucleic acid and polypeptide  
XX sequences at least 80% identical to the TAT nucleic acids and  
XX polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX TAT polypeptide; and methods and compositions for the treatment or  
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX antibodies, antigens, binding molecules and compositions are useful  
XX for diagnosing or treating a cell proliferative disorder associated with  
XX increased TAT expression, particularly cancers such as breast cancer,  
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX used as hybridisation probes, in chromosome and gene mapping, in  
XX chromosome identification and in gene therapy. The present sequence  
XX represents a TAT nucleic acid of the invention

XX Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Query Match 99.8%; Score 1010.4; DB 13; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAGCGCGGTGGAATGGGAGGCGCTGAGAGGAGGTGCTTCAATAGCCTGTGAGC 60  
DB 1000 AGAGCGCGGTGGAATGGGAGGCGCTGAGAGGAGGTGCTTCAATAGCCTGTGAGC 1059

QY 61 CATGTATGAAAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAGACTCTGA 120

DB 1060 CATGTATGAAAGGACCAAGATCTTGATCTAGGCCCTCTGTATGATCCAGATGAGACTCTGA 1119

QY 121 CAACGTGCAATTTATGTAACAAGATTAATGACAGTGTGACTCTGATGATGTGGCAGA 180

DB 1120 CAACGTGCAATTTATGTAACAAGATTAATGACAGTGTGACTCTGATGATGTGGCAGA 1179

QY 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGCAACCCATGATCCA 240

DB 1180 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAAGAACTGGCAACCCATGATCCA 1239

QY 241 CATCTACTGGAACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAGA 300

DB 1240 CATCTACTGGAACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAGA 1299

QY 301 CCCACCCAGTCGCCAAGGCGGTGGAATGTTGATGGGAAAGATTTCAAGGAGCA 360

DB 1300 CCCACCCAGTCGCCAAGGCGGTGGAATGTTGATGGGAAAGATTTCAAGGAGCA 1359

QY 361 ACTTAAGTCTCCCTTCTCTGAGAGAGCCCTCCATGAAACAGTATCGGGGTGTGCTGCC 420

DB 1360 ACTTAAGTCTCCCTTCTCTGAGAGAGCCCTCCATGAAACAGTATCGGGGTGTGCTGCC 1419

QY 421 ACCCGGTGAGGAGGAGGAGCATGCAACCACTCCGTGTGAGGTCTCAGAGGCCCAAGAG 480

DB 1420 ACCCGGTGAGGAGGAGGAGCATGCAACCACTCCGTGTGAGGTCTCAGAGGCCCAAGAG 1479

QY 481 TCCTGGGGGACCCATGAGGTGCGCATGGAGGCGGTGAGAGGATGAGAGGAGCTTCCCTCC 540

DB 1480 TCCTGGGGGACCCATGAGGTGCGCATGGAGGCGGTGAGAGGATGAGAGGAGCTTCCCTCC 1539

QY 541 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAGAAAGTCCAGCACCGAGC 600

DB 1540 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAGAAAGTCCAGCACCGAGC 1599

QY 601 TGGAGACTGGCACTGTCCTCCATTCGGGTTGTGGAACAGAACTTGGCTGAGAGACAGA 660

DB 1600 TGGAGACTGGCACTGTCCTCCATTCGGGTTGTGGAACAGAACTTGGCTGAGAGACAGA 1659

QY 661 GTGCAACAGTGTAAAGGCGCCCAAGGCTGAAGGCTTCTCCGCGCACCCCTTCCGCCCC 720

DB 1660 GTGCAACAGTGTAAAGGCGCCCAAGGCTGAAGGCTTCTCCGCGCACCCCTTCCGCCCC 1719

QY 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGTGGCTCAT 780

DB 1720 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGTGGCTCAT 1779

QY 781 GATCTGTGTGTCCTCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 840

DB 1780 GATCTGTGTGTCCTCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 1839

QY 841 CCGTGTGTCGCGGGGAGATGAGACGAGAGTGGCTTGTGAGAGAAAGAGTGGCCCTGG 900

DB 1840 CCGTGTGTCGCGGGGAGATGAGACGAGAGTGGCTTGTGAGAGAAAGAGTGGCCCTGG 1899

QY 901 GGGGCCCCCTTGACCTTGTATGAGACAGATGGAGAGAAAGAGAGAGTGGAGACC 960

DB 1900 GGGGCCCCCTTGACCTTGTATGAGACAGATGGAGAGAAAGAGAGAGTGGAGACC 1959

QY 961 TGAATAATGATTAAGGCGAGCACTCTCAGAGCCGAGATCGCCCTTAC 1012

DB 1960 TGAATAATGATTAAGGCGAGCACTCTCAGAGCCGAGATCGCCCTTAC 2011

RESULT 11  
ADRO7446  
ID ADRO7446 standard; cDNA; 2026 BP.

AC	ADR07446;
AD	
DT	04-NOV-2004 (first entry)
DE	Full length human cDNA useful for treating neurological disease Seq 952.
XX	
KW	gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KM	osteoporosis; neurological disease; Alzheimer's disease;
KW	Parkinson's disease; dementia; short memory; cancer;
KM	sense or motor function; emotional reaction; fear response; panic;
KW	osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
XX	transqualiser.
OS	Homo sapiens.
XX	
PN	EP1447413-A2.
PD	18-AUG-2004.
XX	
PF	12-FEB-2004; 2004EP-00003145.
XX	
PR	14-FEB-2003; 2003JP-00102207.
XX	
PR	09-MAY-2003; 2003JP-00131452.
XX	
PA	(REMS-) RES ASSOC BIOTECHNOLOGY.
XX	
P1	Isoqal T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,
PI	Wakamatsu A, Ishii S, Nagai K, Irie R,
DR	WPI; 2004-583265/57.
PS	P-PsDB; ADRO9402.
PT	New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
XX	Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX	
PS	Claim 1; SEQ ID NO 952; 2686bp; English.
XX	
CC	This invention relates to novel, isolated full length human cDNA
CC	molecules and the encoded proteins thereof. Specifically, it refers to
CC	cDNA clones obtained by an oligo-capping method, where none of these
CC	clones are identical to any known human mRNAs. The present invention
CC	describes an immunoassay to identify agonists and antagonists, as well as
CC	antibodies, antisense molecules and siRNAs that can all be used to bind
CC	to and modulate expression of the cDNA molecules. As such, these
CC	molecules are useful for diagnostic markers or therapeutic targets for
CC	the various diseases or morbid states. In particular, they are useful in
CC	gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC	disease, Parkinson's disease, dementia, short memory and various cancers,
CC	as well as for maintaining equilibrium of sense or motor function, and
CC	for treating emotional reaction, fear response and panic. Accordingly,
CC	they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC	cyostatic and transqualiser activities. This polynucleotide is a full
CC	length human cDNA sequence of the invention. NOTE: This sequence is not
CC	given in the sequence listing of the specification but can be obtained on
CC	CD-ROM from the European Patent Office, Vienna Sub-office.
XX	
SQ	Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;
XX	
Query Match	99.7%; Score 1008.8; DB 13; Length 2026;
Best Local Similarity	99.8%; Pred.No. 1.7e-23;
Matches 1010; Conservative	0; Mismatches 23; Indels 0; Gaps 0;
OY	1 AGAGCGCGGTGAATGGGCGAGCGCTGAGAGCAGAGTGGCTTCATAAAGCTGGTGACC 60
DB	838 AGAGCGCGGTGAATGGGCGAGCGCTGAGAGCAGAGTGGCTTCATAAAGCTGGTGACC 897
OY	61 CATGATTAAGCACCAAGATTCTTGATCTTAGGCCCACTGTAGATCCAGATGAAGAAGCTCTGA 120
DB	898 CATGATTAAGCACCAAGATTCTTGATCTTAGGCCCACTGTAGATCCAGATGAAGAAGCTCTGA 957
OY	121 CAACGTCGCAATTTATGTACAAGATTAAAGATCAAGCTGTGACTGATGTGATCTGGCAGA 180
DB	958 CAACGTCGCAATTTATGTACAAGATTAAAGATCAAGCTGTGACTGATGTGATCTGGCAGA 1017

QY	181	TTCTCTTAAAGCAGTGTGGGGTTGGTAAGATGAACAAGAACTGGGCAACCCATGATCCA	240
Db	1018	CTTCTTTAAGCAGTGTGGGGTTGGTAAGATGAACAAGAACTGGGCAACCCATGATCCA	1077
QY	241	CATCTACCTGGACAGAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAGA	300
Db	1078	CATCTACCTGGACAGAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAGA	1137
QY	301	CCCAACCACTGSCCAAGCTGCGGTGGAATGTTTGAATGGGAAGATTTTCAAGGACCA	360
Db	1138	CCCAACCACTGSCCAAGCTGCGGTGGAATGTTTGAATGGGAAGATTTTCAAGGACCA	1197
QY	361	ACTTAAAGTCTCCCTTGCTGGAAAGAAAGCCCTCAATGAACAGTATGAGGGGTGCTGCC	420
Db	1198	ACTTAAAGTCTCCCTTGCTGGAAAGAAAGCCCTCAATGAACAGTATGAGGGGTGCTGCC	1257
QY	421	ACCCCTGAGAGGCGAGAGCATATGCCACCACTCTCCGTGAAGGTCCAGAGAGCCAGAGG	480
Db	1258	ACCCCTGAGAGGCGAGAGCATATGCCACCACTCTCCGTGAAGGTCCAGAGAGCCAGAGG	1317
QY	481	TCCTGGGGGAGACCATGGGTGCGATGGAGAGCCGTGGAGAGATGAGAGAGGCTTCCTCC	540
Db	1318	TCCTGGGGGAGACCATGGGTGCGATGGAGAGCCGTGGAGAGATGAGAGAGGCTTCCTCC	1377
QY	541	AAGAGGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGGAAACGTCCACGACGAGC	600
Db	1378	AAGAGGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGGAAACGTCCACGACGAGC	1437
QY	601	TGAGAGCTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCGCTGGAGAACAGA	660
Db	1438	TGAGAGCTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCGCTGGAGAACAGA	1497
QY	661	GTGCAACCACTGTAAAGCCCCCAAGCCTGAAGGCTTCTCCCGCACCTTTCCGCCCC	720
Db	1498	GTGCAACCACTGTAAAGCCCCCAAGCCTGAAGGCTTCTCCCGCACCTTTCCGCCCC	1557
QY	721	GGGTGTGTGATGTGTGGCAGAGGTGSCCCTGTGGGCATCGGGGAGGAAGGTGSCCTCAT	780
Db	1558	GGGTGTGTGATGTGTGGCAGAGGTGSCCCTGTGGGCATCGGGGAGGAAGGTGSCCTCAT	1617
QY	781	GGATCGTGTGTGTCCTCCGTGTGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGGTGACT	840
Db	1618	GGATCGTGTGTGTCCTCCGTGTGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGGTGACT	1677
QY	841	CCGTGTGTGATGTGTGGCAGATCCGAGGTGGCTTTGTGTGAGAGAAAGAGAGTGGCCCTGG	900
Db	1678	CCGTGTGTGATGTGTGGCAGATCCGAGGTGGCTTTGTGTGAGAGAAAGAGAGTGGCCCTGG	1737
QY	901	GGGGCCCCCTTGGACCTTTGATGGAACAGATGGGAGGAAGAGAGAGAGAGAGAGAGAGC	960
Db	1738	GGGGCCCCCTTGGACCTTTGATGGAACAGATGGGAGGAAGAGAGAGAGAGAGAGAGAGC	1797
QY	961	TGAGAAAAATGATTAAGGCGAGACCCGTCAGAGACGAGAGATCGGCTTAC 1012	
Db	1798	TGAGAAAAATGATTAAGGCGAGACCCGTCAGAGACGAGAGATCGGCTTAC 1849	
RESULT 12			
ADP56333	ADP56333 standard; cDNA; 1807 BP.		
XX	AC	ADP56333;	
XX	DT	18-NOV-2004 (first entry)	
XX	DE	Human PRO cDNA sequence SEQ ID NO:2309.	
XX	KW	human; PRO; immune related disease; inflammatory immune response;	
XX	KW	immune response stimulation; antiallergic; antianaemic; antiarthritic;	
XX	KW	antidiabetic; antidiabetic; antinflammatory; antiporiatic;	
XX	KW	antirheumatic; antihypoid; CNS; dermatological; gastrointestinal;	
XX	KW	haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;	

RESULT 12  
ADP56333  
ID ADP56333 standard; cDNA; 1807 BP.

DT	18-NOV-2004	(first entry)
XX		
DE	Human PRO cDNA sequence	SEQ ID NO:2309

KM human, PRO; immune related disease, inflammatory immune response;  
KM immune response stimulation; antiallergic; antianaemic; antiarthritic;  
KM antidiabetic; antidiabetic; antinflammatory; antipsoriatic;  
KM antipneumatic; antithyroid; CNS; dermatological; gastrointestinal;  
KM haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular

KM nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
 KM vitruclide; gene therapy; gene; ss.  
 OS Homo sapiens.  
 XX MO2004039956-A2.  
 XX 13-MAY-2004.  
 XX 28-OCT-2003; 2003WO-US034381.  
 XX 29-OCT-2002; 2002US-0422472P.  
 XX (GETH ) GENENTECH INC.  
 PA Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;  
 PI Wood WI, Wu TD;  
 PI WPI; 2004-376182/35.  
 DR P-PSDB; ADP65334.  
 XX  
 PT New PRO polynucleotides and polypeptides, useful in useful in diagnosing  
 PT and treating an immune related disease, e.g. systemic lupus  
 PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in  
 PT stimulating an immune response.  
 PT  
 PS Claim 2; SEQ ID NO 2309; 3009pp; English.  
 XX  
 CC The present invention describes an isolated PRO nucleic acid (1). Also  
 CC described: (1) a vector comprising (1); (2) a host cell comprising the  
 CC vector of (1); (3) a process for producing a PRO polypeptides; (4) an  
 CC isolated PRO polypeptide; (5) a chimeric molecule comprising the  
 CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
 CC antibody which specifically binds to a polypeptide of (4); (7) a  
 CC composition of matter comprising a polypeptide of (4), an agonist or  
 CC antagonist of the polypeptide or an antibody that binds to the  
 CC polypeptide in combination with a carrier; (8) an article of manufacture  
 CC comprising a container, a label on the container and a composition of  
 CC matter of (7); (9) a method of treating an immune related disease in a  
 CC mammal; (10) a method for determining the presence of a PRO polypeptide  
 CC in a sample suspected of having the polypeptide; (11) a method of  
 CC diagnosing an immune related disease or an inflammatory immune response  
 CC in mammal; (12) a method of identifying a compound that inhibits or  
 CC mimics the activity of or expression of a gene encoding a PRO polypeptide  
 CC; and (13) a method of stimulating the immune response in a mammal. The  
 CC PRO sequences have anti-allergic, anti-inflammatory, antipruritic,  
 CC antiaesthetic, antidiabetic, antitumor, antineoplastic, antipruritic,  
 CC antineoplastic, antitumor, CNS, dermatological, gastrointestinal,  
 CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,  
 CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and  
 CC vitruclide activities and can be used in gene therapy. The nucleic acid  
 CC (1) and the encoded polypeptides, compositions, kits and methods are  
 CC useful in diagnosing and treating an immune related disease and in  
 CC stimulating an immune response. The present sequence represents a human  
 CC PRO nucleotide sequence from the present invention.  
 XX  
 SQ Sequence 1807 BP; 476 A; 497 C; 523 G; 311 T; 0 U; 0 Other:  
 Query Match 94.3%; Score 954.4; DB 13; Length 1807;  
 Best Local Similarity 99.9%; Pred. No. 2e-219;  
 Matches 955; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 237 TCACATCTACCTGAGCAAGAAAGAAAGCCAAAGGCGATGCGACAGTGTCTATG 236  
 DB 1029 TCACATCTACCTGAGCAAGAAAGAAAGCCAAAGGCGATGCGACAGTGTCTATG 1088  
 QY 297 AAGACCCACCACTGCGAAGGCTCCGCGTAATGTTTGAATGGAAGATTTTCAAGGA 356  
 DB 1089 AAGACCCACCACTGCGAAGGCTCCGCGTAATGTTTGAATGGAAGATTTTCAAGGA 1148  
 QY 357 GCAAACTTAAAGTCTCTCTGCTGCGAAGAGCTTCCATGAAACATGATGCGGGGTGCTC 416  
 DB 1149 GCAAACTTAAAGTCTCTCTGCTGCGAAGAGCTTCCATGAAACATGATGCGGGGTGCTC 1208  
 QY 417 TGCCACCCCGTGAAGGCGAAGGATGCAACCACTCCGTGAGGTCAGAGAGCCAG 476  
 DB 1209 TGCCACCCCGTGAAGGCGAAGGATGCAACCACTCCGTGAGGTCAGAGAGCCAG 1268  
 QY 477 GAGTCTTGAGGAGCAATGAGTCCGATGAGAGCCGTGAGAGATAGAGAGGCTTCC 536  
 DB 1269 GAGTCTTGAGGAGCAATGAGTCCGATGAGAGCCGTGAGAGATAGAGAGGCTTCC 1328  
 QY 537 CTCCAAAGAGAGCCCGGGGTTTCCGAGGGAACCCCTCTGAGAGAGAAAGTCCAGAC 596  
 DB 1329 CTCCAAAGAGAGCCCGGGGTTTCCGAGGGAACCCCTCTGAGAGAGAAAGTCCAGAC 1388  
 QY 597 GAGCTGAGAGCTGCGAGTGTCCCAATCCGGGTTTGGAACCAAGAACTTGCTGAGAA 656  
 DB 1389 GAGCTGAGAGCTGCGAGTGTCCCAATCCGGGTTTGGAACCAAGAACTTGCTGAGAA 1448  
 QY 657 CAGAGTGCAACCAAGTAAAGGCCCCAAAGCTTGAAGGCTTCTCCGCAACCTTTCCG 716  
 DB 1449 CAGAGTGCAACCAAGTAAAGGCCCCAAAGCTTGAAGGCTTCTCCGCAACCTTTCCG 1508  
 QY 717 CCCCCGGTGTATCGTGGAGAGTGGCCCTGCTGAGAGAGAGAGAGAGAGAGAGAG 776  
 DB 1509 CCCCCGGTGTATCGTGGAGAGTGGCCCTGCTGAGAGAGAGAGAGAGAGAGAGAG 1568  
 QY 777 TCATGATCGTGTGTGCTCCGAGTGAATGTTCAAGAGTGGCCCTGCTGAGAGAGAGAG 836  
 DB 1569 TCATGATCGTGTGTGCTCCGAGTGAATGTTCAAGAGTGGCCCTGCTGAGAGAGAGAG 1628  
 QY 837 GCTTCGCTGTGTGCGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896  
 DB 1629 GCTTCGCTGTGTGCGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1688  
 QY 897 CTGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAAGAGAGAGAGAGAGAGAGAG 956  
 DB 1689 CTGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAAGAGAGAGAGAGAGAGAGAG 1748  
 QY 957 GACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCGCAAGATCGGCTTAC 1012  
 DB 1749 GACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCGCAAGATCGGCTTAC 1804  
 RESULT 13  
 ID ACN37537  
 ID ACN37537 standard; cDNA; 2177 BP.  
 AC ACN37537;  
 XX 18-NOV-2004 (first entry)  
 DE Tumour-associated antigenic target (TAT) cDNA DNA323971, SEQ ID NO:490.  
 XX  
 XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KM cervical nervous system cancer; bladder cancer; pancreatic cancer;  
 KM cervical cancer; melanoma; leukemia; hybridisation probe;  
 KM chromosome identification; chromosome mapping; gene mapping;  
 KM gene therapy; cytostatic; gene; ss.  
 XX  
 XX Homo sapiens.

```

XX  MO2004030615-A2.
PN
XX
PD  15-APR-2004.
XX
XX  29-SEP-2003; 2003MO-US028547.
XX
PR  02-OCT-2002; 2002US-0414971P.
XX
PA  (GENTH ) GENENTECH INC.
PI  Wu TD, Zhang Z, Zhou Y;
XX
XX  WPI: 2004-347921/32.
DR  P-PSDB; ABM60197.
XX
PT  New tumor-associated antigenic target polypeptides and nucleic acids,
PT  useful in preparing a medicament for treating or detecting a
PT  proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT  prostate cancer or tumor.
XX
PS  Claim 1; SEQ ID NO 490; 7273bp; English.
XX
XX  The invention relates to human tumour-associated antigenic target (TAT)
CC  polypeptides, and their related nucleic acids. The TAT polypeptides are
CC  overexpressed in cancer tissues compared to normal tissues, and may thus
CC  serve as effective targets for the diagnosis and treatment of cancer in
CC  mammals. The invention also relates to nucleic acid and polypeptide
CC  sequences at least 80% identical to the TAT nucleic acids and
CC  polypeptides; expression vectors and host cells comprising a TAT nucleic
CC  acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC  molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC  TAT polypeptide; and methods and compositions for the treatment or
CC  diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC  antibodies, antagonists, binding molecules and compositions are useful
CC  for diagnosing or treating a cell proliferative disorder associated with
CC  increased TAT expression, particularly cancers such as breast cancer,
CC  colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC  cancer, pancreatic cancer, cervical cancer, cancers of the central
CC  nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC  used as hybridisation probes, in chromosome and gene mapping, in
CC  chromosome identification and in gene therapy. The present sequence
CC  represents a TAT nucleic acid of the invention
XX
XX  Sequence 2177 BP; 602 A; 539 C; 605 G; 431 T; 0 U; 0 Other;
SQ
Query Match      89.7%; Score 908; DB 13; Length 2177;
Best Local Similarity 94.9%; Pred. No. 3,1e-208;
Matches 960; Conservative 0; Mismatches 50; Indels 2; Gaps 2;
QY  1 AGGACCGCGGTGGAATGGGACGCGCTGGAGACGAGGTGGCTTCATATAGCCTGGTGACC 60
DB  990 AGGATCGGGTGAATGGGACGCGCTGGAGACGAGGTGGCTTCATATAGCCTGGTGACC 1049
QY  61 CATGATGAGAGACCAATCTTGATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 120
DB  1050 CATGATGAGAGACCAATCTTGATCTAGGCCCACTGTATGATCCAGATGAAGACTCTCTGA 1109
QY  121 CAACAGTGAATTTATGATCAAGAGATTAAATGACAGTGTACTCTGATGATCGGCGAGA 180
DB  1110 CAACAGTGAATTTATGATCAAGAGATTAAATGACATGTATCTCTATATATATCTGGTGA 1169
QY  181 CTTCTTTAAGCACTGTGGGCTTGTATGATGAACAAGAACTGGGCAACCCATGATCCA 240
DB  1170 CTTCTTTAAGCACTGTGGGCTTGTATGATGAACAAGAACTGAGAACCCATGATCCA 1229
QY  241 CATCTACCTGGACAAAGAAACAAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 300
DB  1230 CACCTACCTGGACAAAGAAACAAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 1289
QY  301 CCCACCACTGCGCAAGGCTGCGTGAATGTTTATGATGGAAGATTTTCAAGGGAGCAA 360
DB  1290 CTCACCTACTGCGCAAGGCTGCGTGAATGTTTATGATGGAAGATTTTCAAGGGAGCAA 1349

```

```

QY  361 ACTTAAGTCTCCCTTGTCTCGAAGAAAGCCTCCATGAATGAACAGTATGGCGGGTGTCTGCC 420
DB  1350 ACTTAAGTCTCTCTTGTCTCGAAGAGGCTCTCAATGAACAGTATGAAGGAGTGTATGCC 1409
QY  421 ACCCGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB  1410 ACCCGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
QY  481 TCCTGGGGAGCCCATGAGGTGTGATGAGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 540
DB  1470 TCCTGGGGAGCCCATGAGGTGTGATGAGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 1529
QY  541 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB  1530 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1589
QY  601 TGGAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB  1590 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1649
QY  661 GTGCAACCAAGTGTAAAGGCGCCCAAGGCTTCTCCGCAACCTTTCCGCGCCC 720
DB  1650 GAGCAACCAAGTGTAAAGGCTTCTCCGCAAGGCTTCTCCGCAACCTTTCCGCAACCCC 1709
QY  721 GGGTGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB  1710 GGGTGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1769
QY  781 GGATCGTGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB  1770 GGATCGTGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1829
QY  841 CCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB  1830 CCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
QY  901 GGGGCGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB  1889 GGGGCGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1947
QY  961 TGGAAAAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012
DB  1948 TGGAAAAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1999

```

RESULT 14  
AAS62623/c  
ID AAS62623 standard; cDNA; 2273 BP.

AC AAS62623;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
DE cDNA sequence #410 encoding novel human secreted protein.  
XX  
XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
XX immune deficiency disorder; blood disorder; inflammatory disorder;  
XX infectious disorder; gene therapy; antimicrobial; hepatotropic;  
XX immunosuppressive; antirheumatic; ss.  
XX  
OS Homo sapiens.  
XX  
XX MO200177291-A2.  
XX  
PD 18-OCT-2001.  
XX  
XX 29-MAR-2001; 2001MO-US010485.  
XX  
XX 06-APR-2000; 2000US-0195604P.  
XX  
XX (GENY ) GENETICS INST INC.  
XX  
XX

PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham UR;  
XX MPI; 2002-010900/01.

PT New polynucleotides encoding secreted proteins useful for treating e.g.  
PT asthma, HIV and Crohn's disease.

XX Claim 1; Page 291; 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
XX which encode human secreted proteins. The cDNA sequences have been  
XX derived from a variety of human tissues. The invention also provides a  
XX method for producing proteins from these polynucleotide sequences. The  
XX proteins are useful for identifying compounds that modulate their  
XX activity and production, and the cell is also useful for identifying  
XX compounds that modulate expression of the polynucleotide sequences  
XX encoding the secreted proteins. The sequences of the invention are useful  
XX for treating diseases such as hyperproliferative disorders (e.g. cancer),  
XX immune deficiency disorders (e.g. severe combined immunodeficiency  
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
XX (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
XX infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
XX the invention are also useful in gene therapy. AAs62214-AAs62838  
XX represent the cDNA sequences of the invention that encode for novel human  
XX secreted proteins

XX Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;

Query Match 87.2%; Score 882.4; DB 6; Length 2273;

Best Local Similarity 94.6%; Pred. No. 4.6e-202;

Matches 957; Conservative 0; Mismatches 1; Indels 54; Gaps 2;

QY 1 AGAAGCGGCTGGATGCGCAGCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGACC 60  
DB 1310 AGAAGCGGCTGGATGCGCAGCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGACC 1254  
QY 61 CATGATGAGAGAGCCAGATCTTGATCTAGGCGCCAGCTGTAGTCCAGATGAAGCTCTGA 120  
DB 1253 CATGATGAGAGAGCCAGATCTTGATCTAGGCGCCAGCTGTAGTCCAGATGAAGCTCTGA 1194  
QY 121 CAAAGTGAATTTATGATCAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180  
DB 1193 CAAAGTGAATTTATGATCAAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1134  
QY 181 CTCTTTTAAGCAGTGTGGGTTGTTAAAGTGAACAGAGAACTGGGCAACCCATGATCCA 240  
DB 1133 CTCTTTTAAGCAGTGTGGGTTGTTAAAGTGAACAGAGAACTGGGCAACCCATGATCCA 1074  
QY 241 CATCTACCTGGAAGAGAAAGAGAAAGCCAAAGGAGTCCAGTGTCTATGAGAGA 300  
DB 1073 CATCTACCTGGAAGAGAAAGAGAAAGCCAAAGGAGTCCAGTGTCTATGAGAGA 1014  
QY 301 CCCAGCCACTGCAAGGCTGCGTGAATGTTGATGAGAAAGATTTTCAAGGAGCA 360  
DB 1013 CCCAGCCACTGCAAGGCTGCGTGAATGTTGATGAGAAAGATTTTCAAGGAGCA 954  
QY 361 ACTTAAAGTCTCCTTCTCTGGAAGAAAGCTTCAATGAAGTATGCGGGTGTCTGCC 420  
DB 953 ACTTAAAGTCTCCTTCTCTGGAAGAAAGCTTCAATGAAGTATGCGGGTGTCTGCC 894  
QY 421 ACCCGGAGAGGAG 480  
DB 893 ACCCGGAGAGGAG 834  
QY 481 TCTGAGGAG 540  
DB 833 TCTGAGGAG 774  
QY 541 AAG 600  
DB 773 AAG 714

QY 601 TGGAGACTGAGAGAGTCCCAATCCGGGTTGTGAAACAGAACTTGCCTGGAGAGAGAGA 660  
DB 713 TGGAGACTGAGAGAGTCCCAATCCGGGTTGTGAAACAGAACTTGCCTGGAGAGAGAGA 654  
QY 661 GTGCAACAGTGTAG 720  
DB 653 GTGCAACAGT----- 643  
QY 721 GGGTGTGATTCGTCGAG 780  
DB 642 --GTGATGATTCGTCGAG 585  
QY 781 GATCTGAGTGTGTCCTGTCGATGATTTAGAGAGTGGCCGTGTGAGAGAGAGAGAGAG 840  
DB 584 GATCTGAGTGTGTCCTGTCGATGATTTAGAGAGTGGCCGTGTGAGAGAGAGAGAGAG 525  
QY 841 CCGTGTGAGCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 524 CCGTGTGAGCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465  
QY 901 GGGGCCCCCTTGACCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 464 GGGGCCCCCTTGACCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405  
QY 961 TGGAAATATGATTAAG 1012  
DB 404 TGGAAATATGATTAAG 353

RESULT 15

AB199383  
ID AB199383 standard; cDNA; 2188 BP.

XX AB199383;

XX 07-MAR-2002 (first entry)

XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.

XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

XX Mus musculus.

XX M0200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001MO-JP004192.

XX 18-MAY-2000; 2000JP-00145977.

XX (UNIV-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Iehikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX MPI; 2002-034733/04.

XX P-PSDB; ABB57126.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
XX expression levels of particular genes defined in the specification or by  
XX determining the expression profile of a gene group comprising these

XX genes.

XX Claim 2; Page 794-799; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
XX conditions, comprising measuring the expression levels of particular  
XX genes (I) in a test sample or determining the expression profile of a  
XX gene group in the sample comprising genes selected from (I). The method  
XX is useful for examining the ischaemic condition (e.g. compressive  
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring the  
XX expression levels of particular genes (AB199202 to AB199912, encoding the

CC protein sequences in ABB57020 to ABB57174) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischemic condition-improving drugs or  
CC therapeutics for ischemic diseases. ABB59913 and ABB59914 represent PCR  
CC primers for a mouse ischemic condition related sequence, which are used  
CC in the exemplification of the present invention

XX Sequence 2188 BP; 587 A; 551 C; 638 G; 412 T; 0 U; 0 Other;

Query Match 86.9%; Score 879.2; DB 6; Length 2188;  
Best Local Similarity 91.7%; Pred. No. 2,7e-201;  
Matches 928; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
QY 1 AGGACGCGGTGGAATGGGCGAGCGCTGGAGAGCGAGGTGGCTTCAATAAGCTTGTTGACC 60
DB 1010 AGGAGGACGCGGTGGAATGGGCGCTGGAGAGCGAGGTGGCTTCAATAAGCTTGTTGACC 1069
QY 61 CATGGATGAAGAGACCAATCTTGATCTAAGCCCACTGTAGATCCAGATGAAGACTCTGA 120
DB 1070 CATGGATGAAGAGACCAATCTTGATCTAAGCCCTTCTATAGATCCAGATGAAGACTCTGA 1129
QY 121 CAACAGTGAATTTATATACAGAGATTAATGACAGTGAATCTAGATGATCTGACAGA 180
DB 1130 CAACAGTGAATTTATATGCAAGATTAATGACAGTGAATCTAGATGATCTGACAGA 1189
QY 181 CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGCAACCCATGATCA 240
DB 1190 CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAGAGAACTGGGCAACCCATGATCA 1249
QY 241 CATCTACCTGGAACAAGAAACAAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA 300
DB 1250 TATCTACTGATTAAGAGAGAGAGAAAGCCCTAAAGGGGACGCAAGTGTCTTATGAAGA 1309
QY 301 CCCAGCCACTGCCAAGGCTGCGTGAATGATGATGAGAAAGATTTCAAGGAGCA 360
DB 1310 TCCACCAACTGCAAGGCTGCGTGAATGATGATGAGAAAGATTTCAAGGAGCA 1369
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTTCAATGAACAGTATGCGGGGTGTTCTGCC 420
DB 1370 ACTTAAAGTCTCTTGTCTGGAAGAAAGCTTCAATGAACAGTATGCGGGGATGCTGCC 1429
QY 421 ACCCGGTGAGGAGAGAGCAATGCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAG 480
DB 1430 ACCCTGTGAGGAGAGAGGAGTATGCCACCACTTGTGAGAGTCCGTGAGAGAGG 1489
QY 481 TCCGTGGGAGACCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540
DB 1490 CCGTGGAGAGACCAATGGGTGCAATGGAGGCGGTGAGAGAGCAAGAGGAGGCTTCCCTCC 1549
QY 541 AAGAGGACCCCGGGGTTCCGAGGAGAAAGCTTGTGAGAGAGAAAGCTCCAGCACCGAGC 600
DB 1550 AAGAGGAGCCCGAGAGCTCCAGAGAGAAAGCTTGTGAGAGAGAAAGTCCAGCACCGAGC 1609
QY 601 TGAAGACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGAGAGACAGA 660
DB 1610 TGAAGACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGAGAGACAGA 1669
QY 661 GTGCAACCAAGTGAAGGCGCCCAAGCTGAAGGCTTCTCCGCGCACCTTTCCGAGCCCC 720
DB 1670 ATGCAACCAAGTGAAGGCGCCCTAAGCCGAGAGGCTTCTCCGCGCACCTTTCCAGCTCC 1729
QY 721 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGAGAGAAAGAGTGGCTCAT 780
DB 1730 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGAGAGAAAGAGAGTGGCTCAT 1789
QY 781 GGAATCTGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 840
DB 1790 GGAATCTGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 1849
QY 841 CCGTGTGAGCCGGGAGCAATGACCGAGGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTTGG 900
DB 1850 CCGAGGTGGCCGTGTGAGATGACCGAGGTGGCTTTGTGTGAGAGAAAGAGTGGCTTGG 1909
```

```
QY 901 GGGGCCCCCTGTAACCTTTGATGGAACAGATGGAGAGAGAGAGAGAGAGTGTGAGACC 960
DB 1910 GGGGCCCCCTGTAACCTTTGATGGAACAGATGGAGAGAGAGAGAGAGAGTGTGAGACC 1969
QY 961 TGAAGAAATGATTAAGGCGAGCAGCGTCAAGAGCGGAGAGATCGGCCCTTAC 1012
DB 1970 TGAAGAAATGATTAAGGCGAGCAGCGTCAAGAGCGGAGAGATCGGCCCTTAC 2021
```

Search completed: February 20, 2005, 05:30:50  
Job time : 495.802 secs



***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53 ; Search time 167.828 Seconds  
(without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012  
Sequence: 1 aggcagcggtggaatggca.....agcgagagatcgccctac 1012

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0 %  
Maximum Match 100 %

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1010.4	99.8	2371	2 US-08-343-443B-1	Sequence 1, Appl
2	956	94.5	1785	4 US-09-949-016-5043	Sequence 5043, Ap
3	956	94.5	1785	4 US-09-949-016-5044	Sequence 5044, Ap
4	911.2	90.0	6002	4 US-09-949-016-13696	Sequence 13696, A
5	860	85.0	1783	4 US-09-949-016-13696	Sequence 13696, A
6	338	33.4	411	4 US-09-621-976-13361	Sequence 13361, A
7	258.4	25.5	35784	4 US-09-949-016-16785	Sequence 16785, A
8	258.4	25.5	35784	4 US-09-949-016-16785	Sequence 16785, A
9	213	21.0	601	4 US-09-949-016-176641	Sequence 176641, A
10	213	21.0	601	4 US-09-949-016-176641	Sequence 176641, A
11	164.8	16.3	1939	4 US-09-919-039-322	Sequence 322, App
12	135.8	13.4	601	4 US-09-949-016-176637	Sequence 176637, A
13	135.8	13.4	601	4 US-09-949-016-176638	Sequence 176638, A
14	135.8	13.4	601	4 US-09-949-016-176639	Sequence 176639, A
15	135.8	13.4	601	4 US-09-949-016-176640	Sequence 176640, A
16	111.4	11.0	450	3 US-09-370-838-145	Sequence 145, App
17	111.4	11.0	450	3 US-09-854-133-145	Sequence 145, App
18	74	7.3	114793	4 US-10-148-806-3	Sequence 3, Appl
19	70.2	6.9	215	4 US-09-513-999C-31510	Sequence 31510, A
20	67	6.6	152132	4 US-09-949-016-13845	Sequence 13845, A
21	67	6.6	152132	4 US-09-949-016-13845	Sequence 13845, A
22	66.4	6.6	311	4 US-09-313-294A-5928	Sequence 5928, App
23	64.6	6.4	601	4 US-09-949-016-176640	Sequence 176640, A
24	59.8	5.9	1505	1 US-07-915-246-1	Sequence 1, Appl
25	59.8	5.9	1505	1 US-07-915-246-1	Sequence 1, Appl
26	58.8	5.8	1225	4 US-09-976-594-416	Sequence 416, App
27	58.8	5.8	34230	4 US-09-949-016-12052	Sequence 12052, A

#### ALIGNMENTS

28	58.8	5.8	128470	4 US-09-949-016-13765	Sequence 13765, A
29	58.6	5.8	7218	1 US-08-232-463-14	Sequence 14, Appl
30	57.8	5.7	12695	4 US-09-949-016-16775	Sequence 16775, A
31	57	5.6	1926	3 US-09-249-585A-4	Sequence 4, Appl
32	57	5.6	1931	2 US-09-130-114-2	Sequence 2, Appl
33	55.4	5.5	1505	4 US-09-620-112D-544	Sequence 544, App
34	55.4	5.5	5682	4 US-10-164-595-1	Sequence 1, Appl
35	55.2	5.5	242	4 US-09-354-147C-30	Sequence 30, Appl
36	55.2	5.5	767677	4 US-09-949-016-12147	Sequence 12147, A
37	55.2	5.5	767677	4 US-09-949-016-17361	Sequence 17361, A
38	54.6	5.4	525	4 US-09-640-211A-177	Sequence 177, App
39	54.2	5.4	246	4 US-09-547-693-232	Sequence 232, App
40	54	5.3	1638	4 US-09-759-451-412	Sequence 412, App
41	53.6	5.3	324	4 US-09-547-693-234	Sequence 234, App
42	52.8	5.2	339	4 US-09-270-767-26436	Sequence 26436, A
43	52.8	5.2	3835	4 US-09-270-767-10943	Sequence 10943, A
44	51.4	5.1	472	1 US-08-361-467B-2	Sequence 2, Appl
45	51.4	5.1	472	1 US-08-484-332C-2	Sequence 2, Appl

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougasel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2371 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 25..1992  
US-08-343-443B-1

Query Match 99.8%; Score 1010.4; DB 2; Length 2371;  
Best Local Similarity 99.9%; Pred. No. 2.8e-254;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 AGGACGGGCTGGATGGGCGACCGCTGGAGACGAGGCTTCAATTAAGCTGGTACC 60
DB 981 AGGACGGGCTGGATGGGCGACCGCTGGAGACGAGGCTTCAATTAAGCTGGTACC 1040
QY 61 CATGATGAGAGACACAGATCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACTCTGA 120
DB 1041 CATGATGAGAGACACAGATCTTGATCTAGGCCCTCTGTAGATCCAGATGAAGACTCTGA 1100
QY 121 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA 180
DB 1101 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA 1160
QY 181 CTTCTTTAAGCAGTGTGGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 240
DB 1161 CTTCTTTAAGCAGTGTGGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 1220
QY 241 CATCTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAAGTGTCTATGAAGA 300
DB 1221 CATCTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAAGTGTCTATGAAGA 1280
QY 301 CCCAACCACTGCGCAGAGCTGCGTGGATGTTGATGGGAAAGATTTTCAAGGAGCA 360
DB 1281 CCCAACCACTGCGCAGAGCTGCGTGGATGTTGATGGGAAAGATTTTCAAGGAGCA 1340
QY 361 ACTTAAAGTCTCCCTTGTCTGGAGAAAGCCCTCAATGAACGATATGGGGGTGCTGCC 420
DB 1341 ACTTAAAGTCTCCCTTGTCTGGAGAAAGCCCTCAATGAACGATATGGGGGTGCTGCC 1400
QY 421 ACCCGGTGAGGGGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCGCCAGAGG 480
DB 1401 ACCCGGTGAGGGGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCGCCAGAGG 1460
QY 481 TCTGTGGGGGACCCATGGGTGCGATGGAGGCGCTGTGAGAGATGAGAGGCTTCCCTCC 540
DB 1461 TCTGTGGGGGACCCATGGGTGCGATGGAGGCGCTGTGAGAGATGAGAGGCTTCCCTCC 1520
QY 541 AAGAGGACCCCGGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 600
DB 1521 AAGAGGACCCCGGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 1580
QY 601 TGAAGACTGAGCAGTGTCCCAATTCGAGGTGTGAGAAACAGAACTTCCCTGAGAAACAGA 660
DB 1581 TGAAGACTGAGCAGTGTCCCAATTCGAGGTGTGAGAAACAGAACTTCCCTGAGAAACAGA 1640
QY 661 GTGCAACAGTGTAAAGGCCCAAAAGCTGAAAGGCTTCTCCCGCACCTTTCCGCCCC 720
DB 1641 GTGCAACAGTGTAAAGGCCCAAAAGCTGAAAGGCTTCTCCCGCACCTTTCCGCCCC 1700
QY 721 GGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGCATGCGGGGAGAGAAAGGTGAGCTCAT 780
DB 1701 GGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGCATGCGGGGAGAGAAAGGTGAGCTCAT 1760
QY 781 GGATCTGTGTGTGTCCTCGGTGGAATGTTCAAGGTGCGCCGTGTGAGAGACAGAGGTGCTT 840
DB 1761 GGATCTGTGTGTGTCCTCGGTGGAATGTTCAAGGTGCGCCGTGTGAGAGACAGAGGTGCTT 1820
QY 841 CCGTGTGTGGCCGGGCGATGACCCGAGGTGCTTTGTGTGAGAGAAAGAGAGGTGCGCTGG 900
DB 1821 CCGTGTGTGGCCGGGCGATGACCCGAGGTGCTTTGTGTGAGAGAAAGAGAGGTGCGCTGG 1880
```

```
QY 901 GGGGCCCCCTGACCTTTGATGAGAACAGATGGGAGAGAAAGAGAGACGTGAGGACC 960
DB 1881 GGGGCCCCCTGACCTTTGATGAGAACAGATGGGAGAGAAAGAGAGACGTGAGGACC 1940
QY 961 TGGAAAATATGATTAAGGCGAGACCGGTGAGAGCCGAGAGATGGCCCTTAC 1012
DB 1941 TGGAAAATATGATTAAGGCGAGACCGGTGAGAGCCGAGAGATGGCCCTTAC 1992
```

## RESULT 2

US-09-949-016-5043  
Sequence 5043, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTNER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5043  
LENGTH: 1785  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-5043

Query Match 94.5%; Score 956; DB 4; Length 1785;  
Best Local Similarity 100.0%; Pred. No. 4.2e-240;  
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 57 GACCCATGATGAGAGACCGATCTGATCTAGGCCACCTGTAGATCCAGATGAAGACT 116
DB 827 GACCCATGATGAGAGACCGATCTGATCTAGGCCACCTGTAGATCCAGATGAAGACT 886
QY 117 CTGACAACAGTGCATTTATGTACAGATTTAAATGACAGTGTGACTAGATGTGG 176
DB 887 CTGACAACAGTGCATTTATGTACAGATTTAAATGACAGTGTGACTAGATGTGG 946
QY 177 CAGACTTTTAAAGCAGTGTGGGTTGTTAAGTGAACAAGAACTGGGCAACCCATGA 236
DB 947 CAGACTTTTAAAGCAGTGTGGGTTGTTAAGTGAACAAGAACTGGGCAACCCATGA 1006
QY 237 TCCACATCTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATG 296
DB 1007 TCCACATCTACCTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATG 1066
QY 297 AAGACCAACCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGAAAGATTTTCAAGGA 356
DB 1067 AAGACCAACCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGAAAGATTTTCAAGGA 1126
QY 357 GCAAACTTAAAGTCTCTGCTGCGAAGAAAGCTTCAATGAACAGTATGCGGGGTGTC 416
DB 1127 GCAAACTTAAAGTCTCTGCTGCGAAGAAAGCTTCAATGAACAGTATGCGGGGTGTC 1186
QY 417 TGCACCCCGGTGAGGGGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCGCCAG 476
DB 1187 TGCACCCCGGTGAGGGGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCGCCAG 1246
QY 477 GAGGTCTGTGGGAGACCATGGGTGCAATGGAGGCGGTGAGAGATAGAGAGGCTTCC 536
DB 1247 GAGGTCTGTGGGAGACCATGGGTGCAATGGAGGCGGTGAGAGATAGAGAGGCTTCC 1306
QY 537 CTCGAAGAGAACCCCGGGGTTCCCGAGGAAACCTCTTGAAGAGAGAAAGTCCAGACCC 596
DB 1307 CTCGAAGAGAACCCCGGGGTTCCCGAGGAAACCTCTTGAAGAGAGAAAGTCCAGACCC 1366
```

```
QY 597 GAGCTGAGACATGCGCACTGTCCCAATCCGGGTTGTGAAACAGAACTTGCCCTGAGAA 656
D 1367 GAGCTGAGACATGCGCACTGTCCCAATCCGGGTTGTGAAACAGAACTTGCCCTGAGAA 1426
QY 657 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTCTCCGCGCAACCCCTTCCGC 716
D 1427 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTCTCTCCGCGCAACCCCTTCCGC 1486
QY 717 CCCCAGGTTGATCTGTGCAAGAGTGGCCCTGTGTGCATGTCCGGGAGAAAGAGTGGCC 776
D 1487 CCCCAGGTTGATCTGTGCAAGAGTGGCCCTGTGTGCATGTCCGGGAGAAAGAGTGGCC 1546
QY 777 TCATGATCTGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTG 836
D 1547 TCATGATCTGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTG 1606
QY 837 GCTTCGTTGATGCGCGGGGAGATGGAACGAGAGTGGCTTGTGTGAGAGAAAGAGTGGCC 896
D 1607 GCTTCGTTGATGCGCGGGGAGATGGAACGAGAGTGGCTTGTGTGAGAGAAAGAGTGGCC 1666
QY 897 CTGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGTGGAG 956
D 1667 CTGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGTGGAG 1726
QY 957 GACCTGAAAAATGATTAAGGAGGACACCGTCAAGAGCGGAGATGCGCCCTAC 1012
D 1727 GACCTGAAAAATGATTAAGGAGGACACCGTCAAGAGCGGAGATGCGCCCTAC 1782
```

## RESULT 3

```
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044
```

```
Query Match 94.5%; Score 956; DB 4; Length 1785;
Best Local Similarity 100.0%; Pred. No. 4.2e-240;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 57 GACCCATGATGAAAGCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACT 116
D 827 GACCCATGATGAAAGCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACT 886
QY 117 CTGACAACAGTGCATTTATGATCAAGATTAATGACAGTGTGACTCTAGATGATCTGG 176
D 887 CTGACAACAGTGCATTTATGATCAAGATTAATGACAGTGTGACTCTAGATGATCTGG 946
QY 177 CAGACTCTTTAGCAGTGTGGGTTGTAAGATGAACAAGAACTGGGCAACCCATGA 236
D 947 CAGACTCTTTAGCAGTGTGGGTTGTAAGATGAACAAGAACTGGGCAACCCATGA 1006
QY 237 TCACATCTACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCCATG 296
```

```
D 1007 TCACATCTACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATG 1066
QY 297 AAGACCAACCACTGTGCCAAGGCTGCGGTGGAATGTTGATGGAAGATTTTCAAGGA 356
D 1067 AAGACCAACCACTGTGCCAAGGCTGCGGTGGAATGTTGATGGAAGATTTTCAAGGA 1126
QY 357 GCAACTTAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAAGAACATATCCGGGTGTGTC 416
D 1127 GCAACTTAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAAGAACATATCCGGGTGTGTC 1186
QY 417 TGCACACCCGTGAGGGGAGAGGATGACCAACCACTCCGTGAGAGTCCAGAGAGCCAG 476
D 1187 TGCACACCCGTGAGGGGAGAGGATGACCAACCACTCCGTGAGAGTCCAGAGAGCCAG 1246
QY 477 GAGTCTGTGGGAGACCCATGAGTGCATGGAAGGCGGTGAGAGATAGAGAGCTTCC 536
D 1247 GAGTCTGTGGGAGACCCATGAGTGCATGGAAGGCGGTGAGAGATAGAGAGCTTCC 1306
QY 537 CTCCAAGAGACCCCGGGGTTCCCGAGAGAAACCTTCTGAGAGAGAAAGTCCAGACC 596
D 1307 CTCCAAGAGACCCCGGGGTTCCCGAGAGAAACCTTCTGAGAGAGAAAGTCCAGACC 1366
QY 597 GAGCTGAGACTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTGCCCTGAGAA 656
D 1367 GAGCTGAGACTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTGCCCTGAGAA 1426
QY 657 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTCTCCGCGCAACCCCTTCCGC 716
D 1427 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTCTCTCCGCGCAACCCCTTCCGC 1486
QY 717 CCCCAGGTTGATCTGTGCAAGAGTGGCCCTGTGTGCATGTCCGGGAGAAAGAGTGGCC 776
D 1487 CCCCAGGTTGATCTGTGCAAGAGTGGCCCTGTGTGCATGTCCGGGAGAAAGAGTGGCC 1546
QY 777 TCATGATCTGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTG 836
D 1547 TCATGATCTGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTG 1606
QY 837 GCTTCGTTGATGCGCGGGGAGATGGAACGAGAGTGGCTTGTGTGAGAGAAAGAGTGGAG 896
D 1607 GCTTCGTTGATGCGCGGGGAGATGGAACGAGAGTGGCTTGTGTGAGAGAAAGAGTGGAG 1666
QY 897 CTGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGTGGAG 956
D 1667 CTGGGGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGTGGAG 1726
QY 957 GACCTGAAAAATGATTAAGGAGGACACCGTCAAGAGCGGAGATGCGCCCTAC 1012
D 1727 GACCTGAAAAATGATTAAGGAGGACACCGTCAAGAGCGGAGATGCGCCCTAC 1782
```

## RESULT 4

```
US-09-949-016-13696
; Sequence 13696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13696
; LENGTH: 6002
; TYPE: DNA
```

ORGANISM: Human  
US-09-949-016-13696

Query Match 90.0%; Score 911.2; DB 4; Length 6002;  
Best Local Similarity 95.1%; Pred. No. 3.8e-228;  
Matches 962; Conservative 0; Mismatches 48; Indels 2; Gaps 2;

1 AGAAGCGGTGGTGAATGGGCAAGCCCTGGAGAGCGAGTGGCTTCAATTAAGCTGTGAGACC 60  
2990 AGAATGGGTGGTGAATGGGCAAGCCCTGGAGAGCGAGTGGCTTCAATTAAGCTGTGAGACC 3049  
61 CATGATGAAGAGCAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
3050 CATGATGAAGAGCAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 3109  
121 CAACAGTGAATTTATGTACAAAGATTAAATGAACAGTGTGAATCTAGATGATCTGAGAGA 180  
3110 CAACAGTGAATTTATGTACAAAGATTAAATGAACAGTGTGAATCTAGATGATCTGAGAGA 3169  
181 CTTCTTTAAGCAGTGTGGGTGTTGTAAGATGAACAAAGAACTGGGCAACCAATGATCCA 240  
3170 CTTCTTTAAGCAGTGTGGGTGTTGTAAGATGAACAAAGAACTGGGCAACCAATGATCCA 3229  
241 CATCTACCTGGAACAAGAAACAAGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGA 300  
3230 CACTTACTGTGAACAAGAAACAAGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGA 3289  
301 CCCAACCCTGCAAGGCTGCTGTGAATGTTGATGGAAGAAAGATTTTCAAGGAGCAA 360  
3290 CTCACTTACTGTGAACAAGGCTGCTGTGAATGTTGATGGAAGAAAGATTTTCAAGGAGCAA 3349  
361 ACTTAAAGTCTCTCTGCTGTGAAGAGGCTCCCAATGAACAGTATGGGGGTGTCTGCC 420  
3350 ACTTAAAGTCTCTCTGCTGTGAAGAGGCTCCCAATGAACAGTATGGGGGTGTATGCC 3409  
421 ACCCGTGAAGGAGCAGAGGATGCAACCACTCTGTGAGGTCCAGAGGCGCCAGAGAG 480  
3410 ACCCGTGAAGGAGCAGAGGATGCAACCACTCTGTGAGGTCCAGAGGCGCCAGAGAG 3469  
481 TCCTGGGGGAGCCCATGGGTGTGATGGAAGGCGCTGTGAAGAGATGAGAGGCTTCCCTCC 540  
3470 TCCTGGGGGAGCCCATGGGTGTGATGGAAGGCGCTGTGAAGAGATGAGAGGCTTCCCTCC 3529  
541 AAGAGGAGCCCGGGGTTCCTCGAGGAGAACCTCTGTGAGAGGAGAACCTTCCAGCAACGAGC 600  
3530 AAGAGGAGCCCGGGGTTCCTCGAGGAGAACCTCTGTGAGAGGAGAACCTTCCAGCAACGAGC 3589  
601 TGAAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAAGAACTTCCCTGTGAGAAACA 660  
3590 TGAAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAAGAACTTCCCTGTGAGAAACA 3649  
661 GTGCAACCAAGTGAAGGCGCCCAAGCCTGAAGGCTTCCCTCCGACCTTCCGCCCC 720  
3650 GAGCAACAAGTGAAGGCTTCCCAAGCCTGAAGGCTTCCCTCCGACCTTCCGCCCC 3709  
721 GGGTGTGATGTGTGGCAGAGTGTGGCTGTGAGAGATTCGGGAGAGAGAGTGTGGCTCAT 780  
3710 GGGTGTGATGTGTGGCAGAGTGTGGCTGTGAGAGATTCGGGAGAGAGAGTGTGGCTCAT 3769  
781 GGAATCGTGTGTGTCCGGTGAAGTTCAGAGTGTGGCTGTGAGAGATTCGGGAGAGTGTGCTT 840  
3770 GGAATCGTGTGTGTCCGGTGAAGTTCAGAGTGTGGCTGTGAGAGATTCGGGAGAGTGTGCTT 3829  
841 CCGTGTGTGTGTGGGAGCATGGAACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
3830 CCGTGTGTGTGTGGGAGCATGGAACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3888  
901 GGGGCCCCCTGTGAGCTTGTGAGAAACAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960  
3889 GGGGCCCCCTGTGAGCTTGTGAGAAACAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3947  
961 TGAAGAAATGAGTAAAGGCGAGCAACCGTCAAGAGCGGAGAGATGGGCTTAC 1012

Db 3948 TGAAGAAATGAGTAAAGGCGAGCACTGTCAAGAGCGGAGAGATCAGCCCTTAC 3999

RESULT 5  
US-09-949-016-1954  
; Sequence 1954, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CI.001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 1954  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-1954

Query Match 85.0%; Score 860; DB 4; Length 1783;  
Best Local Similarity 95.1%; Pred. No. 5.5e-215;  
Matches 909; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

57 GACCCATGATGAAGAGCAAGATCTGATCTAGGCGCCACCTGTAGATCCAGATGAAGACT 116  
827 GACCCATGATGAAGAGCAAGATCTGATCTAGGCGCCACCTGTAGATCCAGATGAAGACT 886  
117 CTGACAACAGTGAATTTATGTACAAGATTAATGACAGTGTGATGATGATCTGAG 176  
887 CTGACAACAGTGAATTTATGTACAAGATTAATGACAGTGTGATGATGATGATCTGAG 946  
177 CAGACTTTCTTTAAGCAGTGTGGGTGTTAAGATGAACAAGAACTGGGCAACCCATGA 236  
947 TAGACTTTCTTTAAGCAGTGTGGGTGTTAAGATGAACAAGAACTGGGCAACCCATGA 1006  
237 TCCACATTAACCTGTGACAAGAGAAACAAGGCGCAAGCGGATGCAAGTGTCTATG 296  
1007 TCCACATTAACCTGTGACAAGAGAAACAAGGCGCAAGCGGATGCAAGTGTCTATG 1066  
297 AAGACCAACCACTGCAAGAGGCTGCGGTGAATGTTGATGAGAAAGATTTTCAAGGGA 356  
1067 AAGACCAACCACTGCAAGAGGCTGCGGTGAATGTTGATGAGAAAGATTTTCAAGGGA 1126  
357 GCAAACTTAAAGTCTCTCTGCTGTGAAGAGGCTTCCCAATGAACAGTATCGGGGTGTC 416  
1127 GCAAACTTAAAGTCTCTCTGCTGTGAAGAGGCTTCCCAATGAACAGTATCGGGGTGTC 1186  
417 TGGCAACCCGTGAGGAGCAGAGCAGTCCACCACTCCGTGAGAGTCCAGAGGCGCCAG 476  
1187 TGGCAACCCGTGAGGAGCAGAGCAGTCCACCACTCCGTGAGAGTCCAGAGGCGCCAG 1246  
477 GAGGTCTGTGGGAGCAGATGGGTGCAATGGGAGGCGGTGAGAGAGATGAGAGGCTTCC 536  
1247 GAGGTCTGTGGGAGCAGATGGGTGCAATGGGAGGCGGTGAGAGAGATGAGAGGCTTCC 1306  
537 CTCGAAGAGAGCCCGGGGTTCCTGAGAGGAACTCTGTGAGAGAGAAAGTCCAGCAAC 596  
1307 CTCGAAGAGAGCCCGGGGTTCCTGAGAGGAACTCTGTGAGAGAGAAAGTCCAGCAAC 1366  
597 GAGCTGGAACCTGCACTGTCCCAATCCGGGTGTGAGAAACAAGAACTTGGCTGTGAGAA 656  
1367 GAGCTGGAACCTGCACTGTCCCAATCCGGGTGTGAGAAACAAGAACTTGGCTGTGAGAA 1426  
657 CAGAGTGAACCAAGTGAAGGCGCCCAAGGCTTCCCTCCGCAACCTTTCGCG 716

[illegible]

## RESULT 6

```

US-09-621-976-13361
Sequence 13361, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins
FILE REFERENCE: GENSET.054P2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 13361
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-13361

```

Query Match	33.4%	Score 338	DB 4	Length 411
Best Local Similarity	99.7%	Pred. No. 1.1e-78		
Matches 349	Conservative	0	Mismatches 0	Indels 1
				Gaps 1

Qy	56	GGAGCCCATGATGAAAGAACCAATCTTGAATCTAGGCCCACTCTGATAGATCCAGATGAAGAC	115
Db	62	GGACCCATGGAAGAAAGGACAGATCTTGAATCTAGGCCCACTCTGATAGATCCAGATGAAGAC	121
Qy	116	TCTGCACACAGTGCATTTATGATCAAGAGTTTAAATGACAGTGTGACTCTAGATGATCTG	175
Db	122	TCTGCACACAGTGCATTTATGATCAAGAGTTTAAATGACAGTGTGACTCTGATGATGATCTG	181
Qy	176	GCAGACTTCTTTAAGCAGTGTGGGATTGTTAAGTAAACACAGAGAACTGGGCAACCCATG	235
Db	182	GCAGACTTCTTTAAGCAGTGTGGGATTGTTAAGTAAACACAGAGAACTGGGCAACCCATG	241
Qy	236	ATCCACATCTACTGTGACAAAGAAACAGAAAGCCCAAGGCCGA-TGCCACAGTGTCTTA	294
Db	242	ATCCACATCTACTGTGACAAAGAAACAGAAAGCCCAAGGCCGATTGCCACAGTGTCTTA	301
Qy	295	TGAAGACCAACCACTGCCAAAGGCTGCGGTGGAAATGGTTGATGGGAAAGATTTTCAAG	354
Db	302	TGAAGACCAACCACTGCCAAAGGCTGCGGTGGAAATGGTTGATGGGAAAGATTTTCAAG	361
Qy	355	GAGCAAACTTAAAGTCTCCCTTGCTCGGAAAGAAAGCCCTCCATATACAGTA	404
Db	362	GAGCAAACTTAAAGTCTCCCTTGCTCGGAAAGAAAGCCCTCCATATACAGTA	411

```

RESULT 7
US-09-949-016-16785
; Sequence 16785, Application US/09949016
; Patent No. 6812339
;
GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: fastseq for windows Version 4.0
; SEQ ID NO 16785
; LENGTH: 35784
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16785

```

## Best I

	Matches	268;	Conservative	0;	Mismatches	16;	Indels	0;	Gaps	0;
OY	692	GGCTTCTCCCGCACACCTTTCCGCCCCGGGGGTGAATCGAGGAGAAGTGACCCTG	751							
Dd	33261	GACTGCCTTTGCCCTTCGTATTCTCAACTTAAGTGTGATCTGGCAGAGTGCCCTGT	33320							
OY	752	GGCATGCGGGAGAAAGAGTGCCCTCATAGATCGTGTGTCCCGGTGAAATGTTCA	811							
Dd	33321	GGCATGCGGGAGAAAGAGTGCCCTCATAGATCGTGTGTCCCGGTGAAATGTTCA	33380							
OY	812	GGTAGCCGTGATGAGACAGAGGTGCTTCCGAGTGAGCCGGGAGCATGACCAGAGTGGC	871							
Dd	33381	GGTAGCCGTGATGAGACAGAGGTGCTTCCGAGTGAGCCGGGAGCATGACCAGAGTGGC	33440							
OY	872	TTTTGTGAGGAAGACGAGGTGGCCCTTGAGGGGCCCCCTGACCTTTGATGACAAGATG	931							
Dd	33441	TTTTGTGAGGAAGACGAGGTGGCCCTTGAGGGGCCCCCTGACCTTTGATGACAAGATG	33500							
OY	932	GGAGGAAGAAGAGAGGACGTGAGGACCTGSGAAAAAATGATATA	975							
Dd	33501	GGAGGAAGAAGAAGAGACGTGAGGACCTGSGAAAAAATGATATA	33544							

## RESULT &

```

US-09-949-016,16786
; Sequence 16786, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 16786
; LENGTH: 35784
; TYPE: DNA
; ORGANISM: Human

```





```
QY 2 GGAAGCGGTGGATATGGGAGCGCTGGAGAGCGGTGCTTCAATAGCCGTGGAGCC 61
DB 825 GGAAGCGAGGTGGCATAGGCGGAGAGTACCGTGTGGCTTCAATTAATTTGGGCGCT 884
QY 62 ATGATATGAGGACCAATCTTATCTAGTCTAGGCCCACTGTAGATCCAGATGAAAGTCTGAC 121
DB 885 CGGACCAAGGATCACTCATGATCCGAAAC-----AGATTAATTCAGAC 929
QY 122 AACATGCAATTTATATGTAAGATTAATGACAGTGTACTGATGATCTGATGATCTGATGAC 181
DB 930 AACAAACCAATCTTGTGTAAGGCGCTGGGTGAGATGTAATTAATGATGATCTGATGAT 989
QY 182 TTCTTAAAGAGTGTGGGTGTTTAAATGAAAGAAAGAACTGGGCAACCATATATCAC 241
DB 990 TACTTCAAGAGATGTTGATTTATTAAGCAAAAGAAACGGGACAGCCCATATTAAT 1049
QY 242 ATCTACTGGAACAAGAAACAGAAAGCCCAAGGCGATGCGACAGTCTCTATGAGAGC 301
DB 1050 TTGTACACAGACAGGAGAACTGGCAAGCTGAAAGGAGAGCAAGTCTCTTTGATGAC 1109
QY 302 CCACCACTGCGCAAGGCTGCGGTGAAATGTTGATGAGAAAGATTTCAAGGAGCAAA 361
DB 1110 CCACTTCACTTAAGAGAGTATGATGTTGATGATGATGATGATGATGATGATGATGAT 1169
QY 362 CTTAAAGTCTCCCTGCTCGGAAGAAAGCTCCATGAAAGTATGCGGGGTGCTGCTCCA 421
DB 1170 ATCAAGGTCTCATTTGCT-----ACTCGCGGAGACATTTAAT 1208
QY 422 CCGCGTGAAGGAGAGGAGGATGCGACCACTCCGTGAGAGTCCAGAGGCGCCAGAGT 481
DB 1209 CGGGGTGTGTGCAATGATGCTGCGAGGCGGAGGCGGAGAGGAGGAGGAGGAGGAGGAG 1268
QY 482 CTTGGGAGGAGCCCATGGTTCGATGAGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCA 541
DB 1269 TATGAGGTGTGTGCAAGTGTGAGTGTGAGGCGGAGAGGATTTCCAGTGG----- 1318
QY 542 AGAGAGCCCGGGGTTCCCGAGGAGACCCCTCTGGAAGAGAAACGTCCAGACCGAGCT 601
DB 1319 -----AGGTGTGCGGTGAGAGACAGGAGCGAGCT 1349
QY 602 GGAAGTGTGCAATGCTCCCAATCCGCGTGTGAGAAACAGAACTTCCGCTGAGAGAGAG 661
DB 1350 GGTGATGAGAGTGTCTTAATCCACTGTGAGAAATGAACTTCTTTGAGAGAGATGAA 1409
QY 662 TGAACCAAGTGTAAAGGCGCCCAAGCCTTAAAGGCTTCCCGGCAACCTTCCGCGCC 721
DB 1410 TGAACCAAGTGTAAAGGCGCCCAAGCCTTAAAGGCTTCCCGGCAACCGGTGCTCAC 1469
QY 722 GGTGTGTATGCTGTGCAAGAGTGTGCTGTGTGAGCAATGCGGAGAGAA-----GAGTGGC 775
DB 1470 ATGGGGGTAACTAACGGGAGTATGCTGTGTGAGCAAGAGGCTATGATCGAGGCGGC 1529
QY 776 CTCATGATGCTGTGTGCTCCCGGTGAAATGTTCAAGAGTGGCC---GTGTGAGAGACAGA 832
DB 1530 TACCGGGGCGCGCGGAGACCTGTGAGGCTTCCGAGGGGCGCGGTGTGTGAGAGACA 1589
QY 833 GGTGGCTTCCGTGTGTGC 850
DB 1590 GGTGGCTTGGCCCTGGC 1607

RESULT 12
US-09-949-016-176637
; Sequence 176637, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
```

```
QY 201 TTGTTAAGATGAACAAGAGAACTGGGCAACCCATGATCCCATCTTACTGGAACAAGAAA 260
DB 25 TTGTCTAGATGAACAAGAGAACTGGGCAACCCATGATCCCATCTTACTGGAACAAGAAA 84
QY 261 CAGGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGACCAACCACTGCAAGGCTG 320
DB 85 CAGGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGACCAACCACTGCAAGGCTG 144
QY 321 CCGTGAATGCTTTGATGG 339
DB 145 CCGTGAATGCTTTGATGG 163

RESULT 13
US-09-949-016-176638
; Sequence 176638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 176638
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-176638

Query Match 13.4%; Score 135.8; DB 4; Length 601;
Best Local Similarity 98.6%; Pred. No. 1e-25;
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 201 TTGTTAAGATGAACAAGAGAACTGGGCAACCCATGATCCCATCTTACTGGAACAAGAAA 260
DB 34 TTGTCTAGATGAACAAGAGAACTGGGCAACCCATGATCCCATCTTACTGGAACAAGAAA 93
QY 261 CAGGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGACCAACCACTGCAAGGCTG 320
DB 94 CAGGAAAGCCCAAGGCGATGCAAGTGTCTATGAAGACCAACCACTGCAAGGCTG 153
QY 321 CCGTGAATGCTTTGATGG 339
DB 154 CCGTGAATGCTTTGATGG 172

RESULT 14
US-09-949-016-176638
; Sequence 176638, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 176638
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-176638

Query Match 13.4%; Score 135.8; DB 4; Length 601;
Best Local Similarity 98.6%; Pred. No. 1e-25;
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

US-09-949-016-176660  
; Sequence 176660, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 176660  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-176660

Query Match 13.4%; Score 135.8; DB 4; Length 601;  
Best Local Similarity 98.6%; Pred. No. 1e-25;  
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 201 TTGTTAAGATGAAACAGAGAACTGGGCAACCATGATCCACATCTACCTGGACAGGAAA 260  
DB 34 TTGCTAGATGAAACAGAGAACTGGGCAACCATGATCCACATCTACCTGGACAGGAAA 93  
QY 261 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 320  
DB 94 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 153  
QY 321 CCGTGAATGTTTGATGG 339  
DB 154 CCGTGAATGTTTGATGG 172

RESULT 15  
US-09-949-016-176661  
; Sequence 176661, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 176661  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-176661

Query Match 13.4%; Score 135.8; DB 4; Length 601;  
Best Local Similarity 98.6%; Pred. No. 1e-25;  
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 201 TTGTTAAGATGAAACAGAGAACTGGGCAACCATGATCCACATCTACCTGGACAGGAAA 260  
DB 25 TTGCTAGATGAAACAGAGAACTGGGCAACCATGATCCACATCTACCTGGACAGGAAA 84

QY 261 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 320  
DB 85 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 144  
QY 321 CCGTGAATGTTTGATGG 339  
DB 145 CCGTGAATGTTTGATGG 163

Search completed: February 20, 2005, 16:33:17  
Job time: 177.828 secs



LENGTH: 1988  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-094-749-1074

Query Match 100.0%; Score 1012; DB 17; Length 1988;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-283;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGGACGGGTGGAATGGGACGCGCTGGAGCGAGGTGGCTTCAATTAAGCTTGAGACC 60  
 800 AGGACGGGTGGAATGGGACGCGCTGGAGCGAGGTGGCTTCAATTAAGCTTGAGACC 859  
 61 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 120  
 860 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 919  
 121 CAACAGTGAATTTATATACAGGATTAATAATGACAGTGTACTTATAGATGATCTGGACA 180  
 920 CAACAGTGAATTTATATACAGGATTAATAATGACAGTGTACTTATAGATGATCTGGACA 979  
 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCA 240  
 980 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCA 1039  
 241 CATCTACCTTGACAGAGAAACAGAGAAACCCCAAGGCGATGCCAGTGTCTTATGAAGA 300  
 1040 CATCTACCTTGACAGAGAAACAGAGAAACCCCAAGGCGATGCCAGTGTCTTATGAAGA 1099  
 301 CCCACCACTGCGCAAGGCGCGTGGATGGTGTGATGGGAAAGATTTTCAAGGAGCA 360  
 1100 CCCACCACTGCGCAAGGCGCGTGGATGGTGTGATGGGAAAGATTTTCAAGGAGCA 1159  
 361 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATGCGGGGTGTCTGCC 420  
 1160 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATGCGGGGTGTCTGCC 1219  
 421 ACCCGTGAAGGACAGAGCATGCCACCACTCTCGTGAAGTTCAGAGAGGCCCAAGAG 480  
 1220 ACCCGTGAAGGACAGAGCATGCCACCACTCTCGTGAAGTTCAGAGAGGCCCAAGAG 1279  
 481 TCCGTGGGGGACCCATGGGTGGCATGGAGGCGGTGAAGAGATGAGAGAGGCTTCCCTCC 540  
 1280 TCCGTGGGGGACCCATGGGTGGCATGGAGGCGGTGAAGAGATGAGAGAGGCTTCCCTCC 1339  
 541 AAGAGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600  
 1340 AAGAGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1399  
 601 TGAAGACTGGCAAGTGTCCCAATCCGGGTTTGAAGAACCAAACTTCCGCTGGAAGACGA 660  
 1400 TGAAGACTGGCAAGTGTCCCAATCCGGGTTTGAAGAACCAAACTTCCGCTGGAAGACGA 1459  
 661 GTGCAACCAAGTGAAGGCCCCAAGAGCTGAAGGCTTCTCCGCGCAACCTTTCGCGCCCC 720  
 1460 GTGCAACCAAGTGAAGGCCCCAAGAGCTGAAGGCTTCTCCGCGCAACCTTTCGCGCCCC 1519  
 721 GGGTGTGATCTGGCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCTCAT 780  
 1520 GGGTGTGATCTGGCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCTCAT 1579  
 781 GGATGTGTGTGTCCCGGTGAAGTTCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCT 840  
 1580 GGATGTGTGTGTCCCGGTGAAGTTCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCT 1639  
 841 CCGTGTGTGTGTCCCGGTGAAGTTCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCT 900  
 1640 CCGTGTGTGTGTCCCGGTGAAGTTCAGAGGTGGCCCTGTGTCAGTGGGAGAGAAAGGTGGCT 1699  
 901 GGGGCCCCCTGGAACCTTGAATGAACAGATGGGAGAGAAAGAGAGACCTGAGAGACC 960  
 1700 GGGGCCCCCTGGAACCTTGAATGAACAGATGGGAGAGAAAGAGAGACCTGAGAGACC 1759

QY 961 TGAATAATGATTAAGGACGACCGCTCAGAGCGCAGAGATCGGCCCTAC 1012  
 DB 1760 TGAATAATGATTAAGGACGACCGCTCAGAGCGCAGAGATCGGCCCTAC 1811

## RESULT 2

US-09-822-830A-49/c  
 Sequence 49, Application US/09822830A  
 Patent No. US20020142952A1  
 GENERAL INFORMATION:  
 APPLICANT: Genetics Institute, Inc.  
 APPLICANT: Wong, Gordon G.  
 APPLICANT: Clark, Hilary  
 APPLICANT: Rechele, Kim  
 APPLICANT: Agostino, Michael J.  
 APPLICANT: Howes, Steven H.  
 APPLICANT: Resnick, Richard J.  
 APPLICANT: Gulukota, Kamalakkar  
 APPLICANT: Graham, James R.  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
 FILE REFERENCE: GIN 6402  
 CURRENT APPLICATION NUMBER: US/09/822,830A  
 PRIOR FILING DATE: 2001-03-29  
 PRIOR APPLICATION NUMBER: 60/195,604  
 PRIOR FILING DATE: 2000-04-06  
 NUMBER OF SEQ ID NOS: 631  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 49  
 LENGTH: 2176  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-822-830A-49

Query Match 100.0%; Score 1012; DB 9; Length 2176;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-283;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGGACGGGTGGAATGGGACGCGCTGGAGCGAGGTGGCTTCAATTAAGCTTGAGACC 60  
 1205 AGGACGGGTGGAATGGGACGCGCTGGAGCGAGGTGGCTTCAATTAAGCTTGAGACC 1146  
 61 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 120  
 1145 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 1086  
 121 CAACAGTGAATTTATGATCAAGAGATTAATAATGACAGTGTACTTATAGATATCTGGAGA 180  
 1085 CAACAGTGAATTTATGATCAAGAGATTAATAATGACAGTGTACTTATAGATATCTGGAGA 1026  
 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCA 240  
 1025 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCA 966  
 241 CATCTACCTTGACAGAGAAACAGAGAAACCCCAAGGCGATGCCAGTGTCTTATGAAGA 300  
 965 CATCTACCTTGACAGAGAAACAGAGAAACCCCAAGGCGATGCCAGTGTCTTATGAAGA 906  
 301 CCCACCACTGCGCAAGGCGCGTGGATGGTGTGATGGGAAAGATTTTCAAGGAGCAA 360  
 905 CCCACCACTGCGCAAGGCGCGTGGATGGTGTGATGGGAAAGATTTTCAAGGAGCAA 846  
 361 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATGCGGGGTGTCTGCC 420  
 845 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACAGTATGCGGGGTGTCTGCC 786  
 421 ACCCGTGAAGGACAGAGCATGCCACCACTCTCGTGAAGTTCAGAGAGGCCCAAGAGG 480  
 785 ACCCGTGAAGGACAGAGCATGCCACCACTCTCGTGAAGTTCAGAGAGGCCCAAGAGG 726  
 481 TCCGTGGGGGACCCATGGGTGGCATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
 725 TCCGTGGGGGACCCATGGGTGGCATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 666

QY 541 AGAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 600  
DB 665 AAGAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGAGAGAGAAAGCTCCAGACCGAGC 606  
QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCGCTGAGAGACGA 660  
DB 605 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCGCTGAGAGACGA 546  
QY 661 GTGCAACCACTGTAAAGGCCCCAAAGCCTGAAAGCTTCCTCCGACCCCTTCGCCCC 720  
DB 545 GTGCAACCACTGTAAAGGCCCCAAAGCCTGAAAGCTTCCTCCGACCCCTTCGCCCC 486  
QY 721 GGGTGTGATGTGTGAGAGAGGTGGCCCTGTGAGATCGGGGAGAGAGAGTGGCCCTCAT 780  
DB 485 GGGTGTGATGTGTGAGAGAGGTGGCCCTGTGAGATCGGGGAGAGAGAGTGGCCCTCAT 426  
QY 781 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 425 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 366  
QY 841 CCGTGTGTGTCGGGGCATGACCGAGGTGGCTTTGTGTGAGAGAAAGCAGAGTGGCCCTGG 900  
DB 365 CCGTGTGTGTCGGGGCATGACCGAGGTGGCTTTGTGTGAGAGAAAGCAGAGTGGCCCTGG 306  
QY 901 GGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGATGGAGAGACC 960  
DB 305 GGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGATGGAGAGACC 246  
QY 961 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGCAGAGATGGCCCTAC 1012  
DB 245 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGCAGAGATGGCCCTAC 194

RESULT 3  
US-10-791-017A-1

Sequence 1, Application US/10791017A  
Publication No. US20040197827A1  
GENERAL INFORMATION:  
APPLICANT: JENAPHARM GmbH & Co. KG  
TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances  
FILE REFERENCE: Pat 3684/11  
CURRENT APPLICATION NUMBER: US/10/791,017A  
CURRENT FILING DATE: 2004-03-02  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 2390  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (44)..(2011)  
OTHER INFORMATION: EMS  
US-10-791-017A-1

Query Match 100.0%; Score 1012; DB 18; Length 2390;

Best Local Similarity 100.0%; Pred. No. 1.6e-283;

Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGGTGATGAGGCGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 60  
DB 1000 AGGAGCGGTGATGAGGCGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 1059  
QY 61 CATGATGAGAGACAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
DB 1060 CATGATGAGAGACAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 1119  
QY 121 CAACAGTGCATTTATATACAGAGATTAATGACAGTGTGACTCTAGATATCTGGCAGA 180  
DB 1120 CAACAGTGCATTTATATACAGAGATTAATGACAGTGTGACTCTAGATATCTGGCAGA 1179  
QY 181 CTCTTTAAGCAGTGTGGGTGTGTTAAGATGAACAAGAGAACTGGGCAACCATGATCCA 240

RESULT 4  
US-09-880-107-3769

Sequence 3769, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:

APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-NO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950

DB 1180 CTCTTTAAGCAGTGTGGGTGTGTTAAGATGAACAAGAGAACTGGGCAACCATGATCCA 1239  
QY 241 CATCTACTGTGACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAGA 300  
DB 1240 CATCTACTGTGACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAGA 1299  
QY 301 CCCAACCACTGCCAAGCTGCCCGGTGAATGTTGATGAGAAAGATTTTCAAGGAGACAA 360  
DB 1300 CCCAACCACTGCCAAGCTGCCCGGTGAATGTTGATGAGAAAGATTTTCAAGGAGACAA 1359  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACAGTATGCGGGGTGTGCC 420  
DB 1360 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACAGTATGCGGGGTGTGCC 1419  
QY 421 ACCCGGTGAGAGAGAGAGCATGTCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAG 480  
DB 1420 ACCCGGTGAGAGAGAGAGCATGTCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAG 1479  
QY 481 TCCGTGGGGAGCCCATGGGTGGCATGAGAGGCGGTGAGAGAGATGAGAGAGCTTCCTCC 540  
DB 1480 TCCGTGGGGAGCCCATGGGTGGCATGAGAGGCGGTGAGAGAGATGAGAGAGCTTCCTCC 1539  
QY 541 AAGAGACCCCGGGGTTCCCGAGAGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGAGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1599  
QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTGCTGAGAGACAGA 660  
DB 1600 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTGCTGAGAGACAGA 1659  
QY 661 GTGCAACCACTGTAAAGGCCCCAAAGCCTTGAAGGCTTCCTCCGCCCCCTTCGCCCC 720  
DB 1660 GTGCAACCACTGTAAAGGCCCCAAAGCCTTGAAGGCTTCCTCCGCCCCCTTCGCCCC 1719  
QY 721 GGGTGTGATGTGTGAGAGAGGTGGCCCTGTGAGATCGGGGAGAGAGAGTGGCCCTCAT 780  
DB 1720 GGGTGTGATGTGTGAGAGAGGTGGCCCTGTGAGATCGGGGAGAGAGAGTGGCCCTCAT 1779  
QY 781 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 1780 GATTCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
QY 841 CCGTGTGTGTCGGGGCATGACCGAGGTGGCTTTGTGTGAGAGAAAGCAGAGTGGCCCTGG 900  
DB 1840 CCGTGTGTGTCGGGGCATGACCGAGGTGGCTTTGTGTGAGAGAAAGCAGAGTGGCCCTGG 1899  
QY 901 GGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGATGGAGAGACC 960  
DB 1900 GGGGCCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGATGGAGAGACC 1959  
QY 961 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGCAGAGATGGCCCTAC 1012  
DB 1960 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGCAGAGATGGCCCTAC 2011

SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3769  
LENGTH: 2390  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020412981A1 X66899  
US-09-880-107-3769

Query Match 99.8%; Score 1010.4; DB 9; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 4.6e-283;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGCGGTGGAATGGGCGCGCTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 60  
DB 1000 AGGACCGGTGGATATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 1059  
QY 61 CATGATGAGAGGACCAATCTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120  
DB 1060 CATGATGAGAGGACCAATCTGATCTAGAGCCCTCTGTAGATCCAGATGAAGCTCTGA 1119  
QY 121 CAACAGTGAATTTATGTAACAAGATTAAATGACAGTGTACTCTAGATGATCTGGAGA 180  
DB 1120 CAACAGTGAATTTATGTAACAAGATTAAATGACAGTGTACTCTAGATGATCTGGAGA 1179  
QY 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGACAAAGAACTGGGGCAACCCATGATCCA 240  
DB 1180 CTCTTTAAGCAGTGTGGGTTGTTAAGATGACAAAGAACTGGGGCAACCCATGATCCA 1239  
QY 241 CATCTACCTGGAGAACAGGAAACAGGAAAGCCAAAGGCGATGCCAAGTGTCTATGAAGA 300  
DB 1240 CATCTACCTGGAGAACAGGAAACAGGAAAGCCAAAGGCGATGCCAAGTGTCTATGAAGA 1299  
QY 301 CCCACCACTGCCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTCAGAGGAGCAA 360  
DB 1300 CCCACCACTGCCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTCAGAGGAGCAA 1359  
QY 361 ACTTAAAGTCTCCCTTCTGCTGGAGAAAGCTTCCATGAACATGATGGGGGTGTCTGCC 420  
DB 1360 ACTTAAAGTCTCCCTTCTGCTGGAGAAAGCTTCCATGAACATGATGGGGGTGTCTGCC 1419  
QY 421 ACCCGTGAAGGAGGAGGATGACCAACCACTCCGTGAGAGTCCAGAGGAGCCAGGAGG 480  
DB 1420 ACCCGTGAAGGAGGAGGATGACCAACCACTCCGTGAGAGTCCAGAGGAGCCAGGAGG 1479  
QY 481 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGAGAGATGAGAGAGGCTTCCCTCC 540  
DB 1480 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGAGAGATGAGAGAGGCTTCCCTCC 1539  
QY 541 AAGAGAACCCCGGGGTTCCGAGAGGAAACCTCTCTGAGAGAGAGAAAGTCCAGACCCAGC 600  
DB 1540 AAGAGAACCCCGGGGTTCCGAGAGGAAACCTCTCTGAGAGAGAGAAAGTCCAGACCCAGC 1599  
QY 601 TGAAGACTGGCAGTGTCCCAATCCGGGTTGAGAAACAGAACTTCCCTGAGAGAGAGA 660  
DB 1600 TGAAGACTGGCAGTGTCCCAATCCGGGTTGAGAAACAGAACTTCCCTGAGAGAGAGA 1659  
QY 661 GTGCAACCAAGTGAAGGCGCCCAAAAGCTGAAAGCTTCTCCGACACCTTTCGCCCCC 720  
DB 1660 GTGCAACCAAGTGAAGGCGCCCAAAAGCTGAAAGCTTCTCCGACACCTTTCGCCCCC 1719  
QY 721 GGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGCAATGCGGGAGAGAAAGGTGGCTCAT 780  
DB 1720 GGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGCAATGCGGGAGAGAAAGGTGGCTCAT 1779  
QY 781 GGATCGTGTGATCTCCGGTGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGCTT 840  
DB 1780 GGATCGTGTGATCTCCGGTGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGCTT 1839  
QY 841 CCGTGTGTGAGCCGAGGAGATGAGCCGAGGTGAGCTTTGTGTGAGAGAAACAGAGTGGCCCTTG 900  
DB 1840 CCGTGTGTGAGCCGAGGAGATGAGCCGAGGTGAGCTTTGTGTGAGAGAAACAGAGTGGCCCTTG 1899

QY 901 GGGGCCCCCTGGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGC 960  
DB 1900 GGGGCCCCCTGGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGC 1959  
QY 961 TGAAGAAATGATTAAGGCGAGCAGCTGAGAGAGCGCAGAGATCGGCTTAC 1012  
DB 1960 TGAAGAAATGATTAAGGCGAGCAGCTGAGAGAGCGCAGAGATCGGCTTAC 2011

RESULT 5  
US-09-960-706-1081

Sequence 1081, Application US/0960706  
Publication No. US20030134280A1

GENERAL INFORMATION:  
APPLICANT: Munger, William E.

TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia

FILE REFERENCE: 44921-5029-01US

CURRENT APPLICATION NUMBER: US/09/960,706

PRIOR APPLICATION NUMBER: 60/223,323

PRIOR FILING DATE: 2000-08-07

PRIOR APPLICATION NUMBER: 09/873,319

PRIOR FILING DATE: 2001-06-05

NUMBER OF SEQ ID NOS: 1124

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1081

LENGTH: 2390

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899  
US-09-960-706-1081

Query Match 99.8%; Score 1010.4; DB 10; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 4.6e-283;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGCGGTGGAATGGGCGCGCTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 60  
DB 1000 AGGACCGGTGGATATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCTGGTGACC 1059  
QY 61 CATGATGAGAGGACCAATCTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120  
DB 1060 CATGATGAGAGGACCAATCTGATCTAGAGCCCTCTGTAGATCCAGATGAAGCTCTGA 1119  
QY 121 CAACAGTGAATTTATGTAACAAGATTAAATGACAGTGTACTCTAGATGATCTGGAGA 180  
DB 1120 CAACAGTGAATTTATGTAACAAGATTAAATGACAGTGTACTCTAGATGATCTGGAGA 1179  
QY 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGACAAAGAACTGGGGCAACCCATGATCCA 240  
DB 1180 CTCTTTAAGCAGTGTGGGTTGTTAAGATGACAAAGAACTGGGGCAACCCATGATCCA 1239  
QY 241 CATCTACCTGGAGAACAGGAAACAGGAAAGCCAAAGGCGATGCCAAGTGTCTATGAAGA 300  
DB 1240 CATCTACCTGGAGAACAGGAAACAGGAAAGCCAAAGGCGATGCCAAGTGTCTATGAAGA 1299  
QY 301 CCCACCACTGCCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTCAGAGGAGCAA 360  
DB 1300 CCCACCACTGCCAAGGCTGCGCTGGAATGTTGATGGAAAGATTTCAGAGGAGCAA 1359  
QY 361 ACTTAAAGTCTCCCTTCTGCTGGAGAAAGCTTCCATGAACATGATGGGGGTGTCTGCC 420  
DB 1360 ACTTAAAGTCTCCCTTCTGCTGGAGAAAGCTTCCATGAACATGATGGGGGTGTCTGCC 1419  
QY 421 ACCCGTGAAGGAGGAGGATGACCAACCACTCCGTGAGAGTCCAGAGGAGCCAGGAGG 480  
DB 1420 ACCCGTGAAGGAGGAGGATGACCAACCACTCCGTGAGAGTCCAGAGGAGCCAGGAGG 1479  
QY 481 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGAGAGATGAGAGAGGCTTCCCTCC 540  
DB 1480 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGAGAGATGAGAGAGGCTTCCCTCC 1539

QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGC 1599  
QY 601 TGGAGACTGCACTGTCTCCATCCGGGTGTGAAACCAAGACTTCCCTGAGAAACAGA 660  
DB 1600 TGGAGACTGCACTGTCTCCATCCGGGTGTGAAACCAAGACTTCCCTGAGAAACAGA 1659  
QY 661 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAGCTTCTCCGACCTTTCCGCCCC 720  
DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAGCTTCTCCGACCTTTCCGCCCC 1719  
QY 721 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 780  
DB 1720 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 1779  
QY 781 GGATCGTGTGTCTCCGCTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 1780 GGATCGTGTGTCTCCGCTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
QY 841 CCGTGTGTGCGCGGGGCAATGACCGAGGTGGCTTTGTGAGAGAAACAGAGTGGCCCTG 900  
DB 1840 CCGTGTGTGCGCGGGGCAATGACCGAGGTGGCTTTGTGAGAGAAACAGAGTGGCCCTG 1899  
QY 901 GGGGCCCCCTGTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGTGGAGAC 960  
DB 1900 GGGGCCCCCTGTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGTGGAGAC 1959  
QY 961 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 1012  
DB 1960 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 2011

## RESULT 6

US-09-873-319-717  
Sequence 717, Application US/09873319A  
Publication No. US20030134324A1  
GENERAL INFORMATION:  
APPLICANT: Munger, William E.  
APPLICANT: Kulkarni, Prakash  
APPLICANT: Getzenberg, Robert H.  
APPLICANT: Mega, Iwao  
APPLICANT: Yamamoto, Jun  
TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
FILE REFERENCE: 44921-5029-US  
CURRENT APPLICATION NUMBER: US/09/873,319A  
EARLIER APPLICATION NUMBER: US 60/223,323  
EARLIER FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 755  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 717  
LENGTH: 2390  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899  
US-09-873-319-717

Query Match 99.8%; Score 1010.4; DB 10; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 4.6e-283;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGACCGGCTGATGAGCGCTGAGAGCGAGGTGCTTAATTAAGCTGTGGAC 60  
DB 1000 AGGACCGGCTGATGAGCGCTGAGAGCGAGGTGCTTAATTAAGCTGTGGAC 1059  
QY 61 CATGATGAAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
DB 1060 CATGATGAAGACCAAGATCTTATCTAGGCCCTCTGTAGATCCAGATGAAGACTCTGA 1119

QY 121 CACAGTGAATTTATGTAACAAGATTAAATGAACAGTGTACTGTAGATGATCTGGAGA 180  
DB 1120 CACAGTGAATTTATGTAACAAGATTAAATGAACAGTGTACTGTAGATGATCTGGAGA 1179  
QY 181 CTTCTTTAAGCACTGTGGGTTGTTAAGATGAACAAGAACTGGCCAACTCATATCA 240  
DB 1180 CTTCTTTAAGCACTGTGGGTTGTTAAGATGAACAAGAACTGGCCAACTCATATCA 1239  
QY 241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGCGATCCACAGTGTCTATGAAGA 300  
DB 1240 CATCTACTGGAACAAGAAACAGAAAGCCCAAGCGATCCACAGTGTCTATGAAGA 1299  
QY 301 CCAACCACTGCCAAGGCTGCGTGAATGCTTGAATGGGAAGATTTTCAAGGAGCA 360  
DB 1300 CCAACCACTGCCAAGGCTGCGTGAATGCTTGAATGGGAAGATTTTCAAGGAGCA 1359  
QY 361 ACTTAAAGTCTCCCTGTCTGGAAGAGCTCCAAATGAACAGATGCGGGGTGTCTGCC 420  
DB 1360 ACTTAAAGTCTCCCTGTCTGGAAGAGCTCCAAATGAACAGATGCGGGGTGTCTGCC 1419  
QY 421 ACCCGTGAAGGAGGACATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAG 480  
DB 1420 ACCCGTGAAGGAGGACATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAG 1479  
QY 481 TCCTGGGGGACCCATGATGCTGCAATGGAGGCGGTGAGAGATGAAGAGGCTTCTCC 540  
DB 1480 TCCTGGGGGACCCATGATGCTGCAATGGAGGCGGTGAGAGATGAAGAGGCTTCTCC 1539  
QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1599  
QY 601 TGGAGACTGCACTGTCTCCATCCGGGTGTGAAACCAAGACTTCCCTGAGAAACAGA 660  
DB 1600 TGGAGACTGCACTGTCTCCATCCGGGTGTGAAACCAAGACTTCCCTGAGAAACAGA 1659  
QY 661 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAGCTTCTCCGACCTTTCCGCCCC 720  
DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAGCTTCTCCGACCTTTCCGCCCC 1719  
QY 721 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 780  
DB 1720 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 1779  
QY 781 GGATCGTGTGTCTCCGCTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 1780 GGATCGTGTGTCTCCGCTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
QY 841 CCGTGTGTGCGCGGGGCAATGACCGAGGTGGCTTTGTGAGAGAAACAGAGTGGCCCTG 900  
DB 1840 CCGTGTGTGCGCGGGGCAATGACCGAGGTGGCTTTGTGAGAGAAACAGAGTGGCCCTG 1899  
QY 901 GGGGCCCCCTGTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGTGGAGAC 960  
DB 1900 GGGGCCCCCTGTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGTGGAGAC 1959  
QY 961 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 1012  
DB 1960 TGGAAAAATGATTAAGGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 2011

## RESULT 7

US-09-822-830A-410/c  
Sequence 410, Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.



APPLICANT: Gulukota, Kamalakar  
 APPLICANT: Graham, James R.  
 FILE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
 FILE REFERENCE: GIN 6402  
 CURRENT APPLICATION NUMBER: US/09/822,830A  
 CURRENT FILING DATE: 2001-03-29  
 PRIOR APPLICATION NUMBER: 60/195,604  
 PRIOR FILING DATE: 2000-04-06  
 NUMBER OF SEQ ID NOS: 631  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 410  
 LENGTH: 2273  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-822-830A-410

Query Match 87.2%; Score 882.4; DB 9; Length 2273;  
 Best Local Similarity 94.6%; Pred. No. 7.5e-246;  
 Matches 957; Conservative 0; Mismatches 1; Indels 54; Gaps 2;

QY 1 AGAAGCGGTGGATGGGCAAGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGTGAGACC 60  
 DB 1310 AGAAGCGGTGGATGGGCAAGCGCTGAGAGCGAGGTGGCTTCAATTAAGCTGTGAGACC 1254  
 QY 61 CATGATGAAGACCAAGATCTTGAATCTAGGCGCCAGCTGTAGATCCAGATGAAGACTCTGA 120  
 DB 1253 CATGATGAAGACCAAGATCTTGAATCTAGGCGCCAGCTGTAGATCCAGATGAAGACTCTGA 1194  
 QY 121 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTACTCTAGATATCTGGCAGA 180  
 DB 1193 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTACTCTAGATATCTGGCAGA 1134  
 QY 181 CTCTTTAAAGCAATGGGGTGTGTAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
 DB 1133 CTCTTTAAAGCAATGGGGTGTGTAGATGAACAAGAACTGGGCAACCCATGATCCA 1074  
 QY 241 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGAGGATGCCAAGTGTCTATGAGA 300  
 DB 1073 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGAGGATGCCAAGTGTCTATGAGA 1014  
 QY 301 CCCAACCCTGCAAGGCTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 DB 1013 CCCAACCCTGCAAGGCTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 954  
 QY 361 ACTTTAAAGTCTCCCTGCTGTGGAAGAAAGCTTCCAAATGAAGATGAGGGGTGTCTGCC 420  
 DB 953 ACTTTAAAGTCTCCCTGCTGTGGAAGAAAGCTTCCAAATGAAGATGAGGGGTGTCTGCC 894  
 QY 421 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
 DB 893 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 834  
 QY 481 TCTGTGGGGAACCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
 DB 833 TCTGTGGGGAACCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 774  
 QY 541 AAGAGG 600  
 DB 773 AAGAGG 714  
 QY 601 TGGAGACTGGCAAGTGTCCCAATCCGGGTGTGGAAGAACAGAACTTCCCTGTGAGAAACA 660  
 DB 713 TGGAGACTGGCAAGTGTCCCAATCCGGGTGTGGAAGAACAGAACTTCCCTGTGAGAAACA 654  
 QY 661 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCCCTCCGCAACCTTTCGCCCC 720  
 DB 653 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCCCTCCGCAACCTTTCGCCCC 643  
 QY 721 GGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGATCCGGGAGGAGAAAGATGGCTCAT 780  
 DB 642 --GTGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGAGATCCGGGAGGAGAAAGATGGCTCAT 585  
 QY 781 GGATCTGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGCTT 840

DB 584 GGATCTGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGCTT 525  
 QY 841 CCGTGTGGCCCGGGGATGAGACCGAGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 900  
 DB 524 CCGTGTGGCCCGGGGATGAGACCGAGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 465  
 QY 901 GGGGCCCCCTGAGACTTGTATGGAACAGATGGGAGGAGAAAGAGAGGAGGAGGAGGAGGAGG 960  
 DB 464 GGGGCCCCCTGAGACTTGTATGGAACAGATGGGAGGAGAAAGAGAGGAGGAGGAGGAGGAGG 405  
 QY 961 TGGAAATAATGATTAAGGCGAGACCGTGCAGAGCCGAGAGATCGGCTTAC 1012  
 DB 404 TGGAAATAATGATTAAGGCGAGACCGTGCAGAGGCGGAGAGATCGGCTTAC 353

RESULT 8  
 US-10-198-846-9847  
 Sequence 9847, Application US/10198846  
 Publication No. US2003009974A1  
 GENERAL INFORMATION:  
 APPLICANT: Lillie, James  
 APPLICANT: Xu, Yongyao  
 APPLICANT: Wang, Youzhen  
 APPLICANT: Steilmann, Kathleen  
 TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
 TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 THERAPY OF BREAST CANCER  
 FILE REFERENCE: MRI-049  
 CURRENT APPLICATION NUMBER: US/10/198,846  
 CURRENT FILING DATE: 2002-07-18  
 PRIOR APPLICATION NUMBER: 60/306,220  
 PRIOR FILING DATE: 2001-07-18  
 NUMBER OF SEQ ID NOS: 14084  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9847  
 LENGTH: 2299  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-198-846-9847

Query Match 56.3%; Score 570; DB 14; Length 2299;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-155;  
 Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GGACCATGATGAAAGACCAAGTCTTGAATCTAGGCGCCACCTGTATGATCCAGTGAAGAC 115  
 DB 1072 GGACCATGATGAAAGACCAAGTCTTGAATCTAGGCGCCACCTGTATGATCCAGTGAAGAC 1131  
 QY 116 TCTGACAAAGTCAATTTATGTAACAAGATTAAATGACAGTGTGACTAGATGATCTG 175  
 DB 1132 TCTGACAAAGTCAATTTATGTAACAAGATTAAATGACAGTGTGACTAGATGATCTG 1191  
 QY 176 GCAGACTTCTTAAGCAAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATG 235  
 DB 1192 GCAGACTTCTTAAGCAAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATG 1251  
 QY 236 ATCAATCTAATCTGGAACAAGAAACAGAAAGCCCAAGAGGATGCCACAGTGTCTAT 295  
 DB 1252 ATCAATCTAATCTGGAACAAGAAACAGAAAGCCCAAGAGGATGCCACAGTGTCTAT 1311  
 QY 296 GAGAGCCCAACCACTGCAAGGCTGCGTGAATGGTTGATGGGAAAGATTTTCAAGGG 355  
 DB 1312 GAGAGCCCAACCACTGCAAGGCTGCGTGAATGGTTGATGGGAAAGATTTTCAAGGG 1371  
 QY 356 AGCAAACTTAAGTCTCCCTGTGCGAAAGAGCTTCAATGAACAGTATGCGGGTGT 415  
 DB 1372 AGCAAACTTAAGTCTCCCTGTGCGAAAGAGCTTCAATGAACAGTATGCGGGTGT 1431  
 QY 416 CTGCAACCCCGTGAAGGAGGAGGATGCCACCACTCGTGTGAGAGTCCAGAGAGGCCA 475  
 DB 1432 CTGCAACCCCGTGAAGGAGGAGGATGCCACCACTCGTGTGAGAGTCCAGAGAGGCCA 1491

Qy 476 GGAGGCTCTGGGGGACCCCATGGTGCATGGGAGGCGGTGGAGGAGATAGAGAGGCTTC 535  
Db 1492 GGAGGCTCTGGGGGACCCCATGGTGCATGGGAGGCGGTGGAGGAGATAGAGAGGCTTC 1551  
Qy 536 CTTCCAAAGAGACCCCGGGGTTCCCGAGGAGAACCTCTGAGAGAGAGAAAGTCCAGACAC 595  
Db 1552 CTTCCAAAGAGACCCCGGGGTTCCCGAGGAGAACCTCTGAGAGAGAGAAAGTCCAGACAC 1611  
Qy 596 CGAGCTGGAGACTGGCTGTCTCCCATCCG 625  
Db 1612 CGAGCTGGAGACTGGCTGTCTCCCATCCG 1641

## RESULT 9

US-09-918-995-27690  
; Sequence 27690, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hvaeq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 27690  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(550)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27690

Query Match 50.2%; Score 508.4; DB 10; Length 550;  
Best Local Similarity 97.5%; Pred. No. 2.6e-137;  
Matches 515; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 335 GATGGGAAGATTTTCAAGGAGCAATTAATCTCCCTTGGTGGAGAGAGCCCTCCA 394  
Db 22 GAAGAGGCAATTTCTTCAAGAGCAATTAATCTCCCTTGGTGGAGAGAGCCCTCCA 81  
Qy 395 ATGAACAGTATGCGGGTGTCTGCAACCCGCTGAGGGAGAGGATGCCACCACTC 454  
Db 82 ATGAACAGTATGCGGGTGTCTGCAACCCGCTGAGGGAGAGGATGCCACCACTC 141  
Qy 455 CGTGAAGTCCAGAGGCGCCAGAGAGTCTTGGGGAGCCATGGGTGCGATGGAGGCGT 514  
Db 142 CGTGAAGTCCAGAGGCGCCAGAGAGTCTTGGGGAGCCATGGGTGCGATGGAGGCGT 201  
Qy 515 GAGAGAGTGAAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCCGAGAGAACCTCTT 574  
Db 202 GAGAGAGTGAAGAGGCTTCCCTCCAAAGAGAACCCCGGGGTTCCCGAGAGAACCTCTT 261  
Qy 575 GAGAGAGAAAGCTCCAGCAACGAGCTGAGAGCTGGAGTGTCCCAATCCGGGTTGGA 634  
Db 262 GAGAGAGAAAGCTCCAGCAACGAGCTGAGAGCTGGAGTGTCCCAATCCGGGTTGGA 321  
Qy 635 AACCAAGACTTCCCTGGAGAAAGAGTGAACCAAGTGAAGGCCCAAGGCTGAAGGC 694  
Db 322 AACCAAGACTTCCCTGGAGAAAGAGTGAACCAAGTGAAGGCCCAAGGCTGAAGGC 381  
Qy 695 TTCTCTCCCGCACCCTTTCCGCGCCCGGGGTGTATGTGTGCAAGAGTGGCCCTGTGGC 754  
Db 382 TTCTCTCCCGCACCCTTTCCGCGCCCGGGGTGTATGTGTGCAAGAGTGGCCCTGTGGC 441  
Qy 755 ATCCGGGAGAGAGAGGTGCTCATGATCGTGTGGTCCCGGTGAAGTTCAGAGGT 814  
Db 442 ATCCGGGAGAGAGAGGTGCTCATGATCGTGTGGTCCCGGTGAAGTTCAGAGGT 501

Qy 815 GGCGGTGTGAGACAGAGGTGCTTCCGTGTGGCCCGGGCATGAC 862  
Db 502 GGCGGTGTGAGACAGAGGTGCTTCCGTGTGGCCCGGGCATGAC 549

## RESULT 10

US-09-918-995-9557  
; Sequence 9557, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hvaeq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 9557  
; LENGTH: 568  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(568)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-9557

Query Match 43.2%; Score 437.2; DB 10; Length 568;  
Best Local Similarity 97.5%; Pred. No. 1.3e-116;  
Matches 466; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

Qy 1 AGGACGGGTGGAATGGGAGCGCTGGAGAGCGAGGTGCTCAATAGCTGGTGGAC 60  
Db 94 AGGACGGGTGGAATGG--GCGCTGAGAGCGAGGTGCTCAATAGCTGGTGGAC 150  
Qy 61 CATGATGAGAGACAGATCTTATCTAGGCGCCAGCTGTAGATCCAGATGAAGCTCTGA 120  
Db 151 CATGATGAGAGACAGATCTTATCTAGGCGCCAGCTGTAGATCCAGATGAAGCTCTGA 210  
Qy 121 CAAAGTGCATTTATGTACAGAGATTAATGAAGTGTACTCTAGATATCTGGAGA 180  
Db 211 CAAAGTGCATTTATGTACAGAGATTAATGAAGTGTACTCTAGATATCTGGAGA 270  
Qy 181 CTCTTTAGCATGTGGGTTGTAAAGAGAAAGAGAACTGGGCAACCATGATCCA 240  
Db 271 CTCTTTAGCATGTGGGTTGTAAAGAGAAAGAGAACTGGGCAACCATGATCCA 330  
Qy 241 CATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCAGATGCTCTATGAGA 300  
Db 331 CATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCAGATGCTCTATGAGA 390  
Qy 301 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 360  
Db 391 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGAAAGATTTTCAAGGAGCAA 450  
Qy 361 ACTTAAAGTCTCCCTGCTGGAAGAAAGCTCCAAAGAAAGATGCGGGGTGTCTGCC 420  
Db 451 ACTTAAAGTCTCCCTGCTGGAAGAAAGCTCCAAAGAAAGATGCGGGGTGTCTGCC 509  
Qy 421 ACCCGTGAAGGAGAGAGCTGCAACCACTCCGTGAAGTCCAGAGAGGCCAGGA 478  
Db 510 ACCCGTGAAGGAGAGAGCTGCAACCACTTGTGAGTCCAGAGAGGCCAGGA 567

## RESULT 11

US-09-925-301-669  
; Sequence 669, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:

```

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 669
LENGTH: 545
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (11)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (13)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (58)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (337)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (453)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-669

```

Query Match 30.5%; Score 309; DB 9; Length 545;  
 Best Local Similarity 78.9%; Pred. No. 2,5e-79;

Matches 396; Conservative 0; Mismatches 93; Indels 13; Gaps 2;

```

QY 467 GGAGGCCAGAGGCTCTGGGGAGCCCATGGTGCATGGAGGCCCTGAGGAGATAGA 526
DB 57 GNCGGCGCTCTAGAGTGAATGATCCCGGCTGAGAGAAATTCGACAGAGATAGA 116
QY 527 GGAGGCTTCCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 586
DB 117 GGAGGCTTCCCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 176
QY 587 GTCCAGACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 646
DB 177 GTCCAGACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 176
QY 647 GCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 706
DB 225 GATTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 284
QY 707 CCGTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766
DB 285 ACAGAGTGCAACAGAGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 344
QY 767 AGAGGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 826
DB 345 AGAGGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 404
QY 827 GACAGAGGTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 886
DB 405 GACAGAGGTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 464
QY 887 CGAGGTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946
DB 465 CGAGGTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523
QY 947 GAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 968

```

DB 524 GGAGGTGGAG 545

RESULT 12

```

US-09-960-352-3851
Sequence 3851, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathiasagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 3851
LENGTH: 361
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 17-LIB3057-024-Q1-K1-E1
US-09-960-352-3851

```

Query Match 26.7%; Score 269.8; DB 9; Length 361;  
 Best Local Similarity 84.2%; Pred. No. 5.5e-68;

Matches 304; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

QY 197 GGAGGTGGTAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 256
DB 1 GGAGGTGGTAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 257 GAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 316
DB 61 TAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 317 GCTGCGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 376
DB 121 GCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 377 GCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 436
DB 181 GCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 437 GGCATGCCACACCACTCCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 496
DB 241 GGCATGCCACACCACTCCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 497 GGTGCGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 556
DB 301 GGTGCGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 557 T 557
DB 361 T 361

```

RESULT 13

```

US-09-864-761-8728
Sequence 8728, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23

```



```
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18163
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR22.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: SWISSPROT HIT: Q07050, EVALUE 2.00e+00
; OTHER INFORMATION: NT HIT: X73003.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE897260.1, EVALUE 0.00e+00
; US-10-029-386-18163

Query Match      25.1%; Score 254.2; DB 16; Length 262;
Best Local Similarity 98.8%; Pred. No. 1,7e-63;
Matches 256; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 717 CCCCAGGTGATCGTGGCAAGGTGCCCCGTGGCATGCGGGAGAGAGGTGCC 776
    |||||||
DB 262 CCTTAGGTGATCGTGGCAAGGTGCCCCGTGGCATGCGGGAGAGAGGTGCC 203
    |||||||

QY 777 TCATGATCGTGGTCCCGGTGGAATGTTCAAGGTGCGGTGGAACAAGGTG 836
    |||||||
DB 202 TCATGATCGTGGTCCCGGTGGAATGTTCAAGGTGCGGTGGAACAAGGTG 143
    |||||||

QY 837 GCTTCCGTGGTGGCCGGGGCATGACCAGGTGCTTTGGTGAAGAAGAGGTGCC 896
    |||||||
DB 142 GCTTCCGTGGTGGCCGGGGCATGACCAGGTGCTTTGGTGAAGAAGAGGTGCC 83
    |||||||

QY 897 CTGGGGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAGAGAGAGCTGAG 956
    |||||||
DB 82 CTGGGGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAGAGAGAGCTGAG 23
    |||||||

QY 957 GACCTGAGAAAATGATTA 975
    |||||||
DB 22 GACCTGAGAAAATGATTA 4
    |||||||
```

Search completed: February 20, 2005, 22:47:58  
Job time : 534.573 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:03:23 ; Search time 2989.19 Seconds  
(without alignments)  
12886.762 Million cell updates/sec

Title: US-10-791-017A-1\_COPY\_1000\_2011

Sequence: 1 aggacgcggtggaatgggca.....aggcagagatcgccctac 1012

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1012	100.0	2073	3	CR593724	CR593724 full-length
2	1012	100.0	2083	3	CR608666	CR608666 full-length
3	1012	100.0	2103	3	CR619493	CR619493 full-length
4	1012	100.0	2157	3	CR604952	CR604952 full-length
5	1012	100.0	2314	3	CR608375	CR608375 full-length
6	996	98.4	2070	3	CR602561	CR602561 full-length
7	996	98.4	2093	3	CR594868	CR594868 full-length
8	996	98.4	2126	3	CR610888	CR610888 full-length
9	996	98.4	2159	3	CR625247	CR625247 full-length
10	892	88.1	2236	3	CR604639	CR604639 full-length
11	882.6	87.2	1379	3	AK014366	AK014366 Mus muscu
12	879.2	86.9	2107	3	AK019460	AK019460 Mus muscu
13	876	86.6	2373	3	AK034755	AK034755 Mus muscu
14	834.8	82.5	1020	5	BX387336	BX387336 BX387336
15	792.8	78.3	2269	5	AK049743	AK049743 Mus muscu
16	789.2	78.0	895	5	BX408406	BX408406 BX408406
17	789.2	78.0	1057	5	BX353312	BX353312 BX353312
18	789.4	78.0	929	5	BO881305	BO881305 AGENCOURT
19	784	77.5	1068	5	BX449773	BX449773 BX449773
20	778	76.9	1048	5	BX381313	BX381313 BX381313
21	777	76.8	962	5	BX407870	BX407870 BX407870
22	770.6	76.1	1013	5	BX465099	BX465099 BX465099
23	769	76.0	1028	5	BX399069	BX399069 BX399069
24	754.4	74.5	838	4	BG574437	BG574437 602596384

25	753.4	74.4	778	4	BG756215	BG756215 602713525
26	751	74.2	1064	5	EX417171	EX417171 BX417171
27	748.6	74.0	1062	5	BX443495	BX443495 BX443495
28	745.6	73.7	1055	5	BX428311	BX428311 BX428311
29	741.2	73.2	792	5	BX368129	BX368129 BX368129
30	740.6	73.2	933	5	BX387305	BX387305 BX387305
31	739.6	73.1	1022	5	BX407825	BX407825 BX407825
32	737.4	72.9	925	5	BX339942	BX339942 BX339942
33	733.8	72.5	964	5	BX382192	BX382192 BX382192
34	732.6	72.4	837	4	BG751179	BG751179 602729679
35	732.4	72.4	1038	5	BX335119	BX335119 BX335119
36	730.2	72.2	901	5	BX347606	BX347606 BX347606
37	722.6	71.7	920	5	BE562296	BE562296 601344870
38	715.4	70.7	787	2	BE741903	BE741903 601594075
39	712.4	70.4	928	5	BX346753	BX346753 BX346753
40	711.4	70.3	897	6	BY713354	BY713354 BY713354
41	701.4	69.3	726	4	BG754333	BG754333 602709891
42	699.6	69.1	951	5	BO642355	BO642355 AGENCOURT
43	697	68.9	849	2	BE563744	BE563744 601335129
44	694.6	68.6	843	7	CN162640	CN162640 952042 MA
45	691.8	68.4	805	5	BX449612	BX449612 BX449612

#### ALIGNMENTS

RESULT 1  
LOCUS CR593724  
DEFINITION full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR593724.1 GI:50474531  
VERSION CR593724.1  
KEYWORDS HTC; CNS/LT; CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1. (bases 1 to 2073)  
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
2 (bases 1 to 2073)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web: www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
FEATURES  
source 1. 2073  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1011Y005"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 100.0%; Score 1012; DB 3; Length 2073;  
Best Local Similarity 100.0%; Pred. No. 86-241;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AGGACGCGGTGGAATGGGCGGCTGAGAGCGAGGCGGTTCAATTAAGCGTGTGAGCC 60  
Db 908 AGGACGCGGTGGAATGGGCGGCTGAGAGCGAGGCGGTTCAATTAAGCGTGTGAGCC 967

QY 61 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 120  
 DB 968 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 1027  
 QY 121 CAACAGTGAATTTATGTAACAAGATTAATGAAGATGATCTGTATGATCTGTGACGA 180  
 DB 1028 CAACAGTGAATTTATGTAACAAGATTAATGAAGATGATCTGTATGATCTGTGACGA 1087  
 QY 181 CTTCTTTAAGCACTGTGGGTTGTAAATGAACAAGAACTGGGCAACCAATGATCCA 240  
 DB 1088 CTTCTTTAAGCACTGTGGGTTGTAAATGAACAAGAACTGGGCAACCAATGATCCA 1147  
 QY 241 CATCTACTCTGACAAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAGA 300  
 DB 1148 CATCTACTCTGACAAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAGA 1207  
 QY 301 CCCAACCCTGACCAAGGCTGCCGTGAAATGATTTGATGGAAAGATTTTCAAGGAGCAA 360  
 DB 1208 CCCAACCCTGACCAAGGCTGCCGTGAAATGATTTGATGGAAAGATTTTCAAGGAGCAA 1267  
 QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAACAGTATGCGGGGTGCTGCC 420  
 DB 1268 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAACAGTATGCGGGGTGCTGCC 1327  
 QY 421 ACCCCGTGAGGGGACAGAGGATGCCACCACTCCGTGAGGTCACAGAGGCCACAGAGG 480  
 DB 1328 ACCCCGTGAGGGGACAGAGGATGCCACCACTCCGTGAGGTCACAGAGGCCACAGAGG 1387  
 QY 481 TCCCTGGGGGACCCATGGGTGACATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 540  
 DB 1388 TCCCTGGGGGACCCATGGGTGACATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 1447  
 QY 541 AAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600  
 DB 1448 AAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1507  
 QY 601 TGAAGATCTGACAGTGTCCCAATCCGGGTGTGGAACAACAACATCTCCCTGTGAGAAACA 660  
 DB 1508 TGAAGATCTGACAGTGTCCCAATCCGGGTGTGGAACAACAACATCTCCCTGTGAGAAACA 1567  
 QY 661 GTGCAACCAAGTGTAAAGGCCCAAGACCTGAAGGCTTCTCCCGCACCTTTCGCGCCCC 720  
 DB 1568 GTGCAACCAAGTGTAAAGGCCCAAGACCTGAAGGCTTCTCCCGCACCTTTCGCGCCCC 1627  
 QY 721 GGGTGTGATCTGTGACAGAGGTGACCTGTGTGAGTCCGGGAGAAAGAGGTGCTCAT 780  
 DB 1628 GGGTGTGATCTGTGACAGAGGTGACCTGTGTGAGTCCGGGAGAAAGAGGTGCTCAT 1687  
 QY 781 GGAATGTGTGTGTCCCGTGAATGTTCAAGAGTGGCCGTGTGAGAACAGAGGTGCTT 840  
 DB 1688 GGAATGTGTGTGTCCCGTGAATGTTCAAGAGTGGCCGTGTGAGAACAGAGGTGCTT 1747  
 QY 841 CCGTGTGTGACCGGAGGACAGACCGAGGTGCTTGTGTGAGAGAGACAGAGGTGCTTGG 900  
 DB 1748 CCGTGTGTGACCGGAGGACAGACCGAGGTGCTTGTGTGAGAGAGACAGAGGTGCTTGG 1807  
 QY 901 GGGGCCCCCTGTGACCTTTGATGAACAAGATGGAAGAAAGAGAGAGACCTGTGAGAAC 960  
 DB 1808 GGGGCCCCCTGTGACCTTTGATGAACAAGATGGAAGAAAGAGAGAGACCTGTGAGAAC 1867  
 QY 961 TGAAGAAATGATTAAGGAGACACCTGTGAGAGAGAGAGATGGGCTTTC 1012  
 DB 1868 TGAAGAAATGATTAAGGAGACACCTGTGAGAGAGAGAGATGGGCTTTC 1919

RESULT 2  
 LOCUS CRE08666 2083 bp mRNA linear HTC 21-JUL-2004  
 DEFINITION full-length cDNA clone CSODG004YL23 of B cells (Ramos cell line) of  
 Homo sapiens (human).  
 ACCESSION CRE08666  
 VERSION CRE08666.1 GI:50489473

KEYWORDS HTC; CNSLT cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2083)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@life.uchicago.edu  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Paradies Avenue  
 Carlsbad, CA 92008  
 REFERENCE 2 (bases 1 to 2083)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES  
 source  
 1..2083  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODG004YL23"  
 /issue\_type="B cells (Ramos cell line)"  
 /plasmid="pCMVSPORT\_6"

ORIGIN  
 Query Match 100.0%; Score 1012; DB 3; Length 2083;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-241;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGGGTGAATGGAGGAGGCTGTGAGAGGAGGCTTCAATAGGCTGTGAGACC 60  
 DB 908 AGGAGCGGGTGAATGGAGGAGGCTGTGAGAGGAGGCTTCAATAGGCTGTGAGACC 967  
 QY 61 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 120  
 DB 908 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 1027  
 QY 968 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAGACTCTGA 1027  
 DB 121 CAACAGTGAATTTATGTAACAAGATTAATGAAGATGATCTGTATGATCTGTGACGA 180  
 QY 121 CAACAGTGAATTTATGTAACAAGATTAATGAAGATGATCTGTATGATCTGTGACGA 180  
 DB 1028 CAACAGTGAATTTATGTAACAAGATTAATGAAGATGATCTGTATGATCTGTGACGA 1087  
 QY 181 CTTCTTTAAGCACTGTGGGTTGTAAATGAACAAGAACTGGGCAACCAATGATCCA 240  
 DB 1088 CTTCTTTAAGCACTGTGGGTTGTAAATGAACAAGAACTGGGCAACCAATGATCCA 1147  
 QY 241 CATCTACTCTGACAAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAGA 300  
 DB 1148 CATCTACTCTGACAAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAGA 1207  
 QY 301 CCCAACCCTGACCAAGGCTGCCGTGAAATGATTTGATGGAAAGATTTTCAAGGAGCAA 360  
 DB 1208 CCCAACCCTGACCAAGGCTGCCGTGAAATGATTTGATGGAAAGATTTTCAAGGAGCAA 1267  
 QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAACAGTATGCGGGGTGCTGCC 420  
 DB 1268 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAACAGTATGCGGGGTGCTGCC 1327  
 QY 421 ACCCCGTGAGGGGACAGAGGATGCCACCACTCCGTGAGGTCACAGAGGCCACAGAGG 480  
 DB 1328 ACCCCGTGAGGGGACAGAGGATGCCACCACTCCGTGAGGTCACAGAGGCCACAGAGG 1387  
 QY 481 TCCCTGGGGGACCCATGGGTGACATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 540  
 DB 1388 TCCCTGGGGGACCCATGGGTGACATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 1447





RESULT 4  
LOCUS CR604952 2157 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D1010Y112 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR604952  
VERSION CR604952.1 GI:50485759  
KEYWORDS HTC; CNSLT\_CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2157)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2157)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source location/Qualifiers  
1..2157  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1010Y112"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 100.0%; Score 1012; DB 3; Length 2157;  
Best Local Similarity 100.0%; Pred. No. 8.1e-241;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGACCGCGGTGGATGGGACGCTGGAGAGCGAGGTGCTTCAATAGCTGTGGACC 60  
DB 997 AGGACGGCGGTGGATGGGACGCTGGAGAGCGAGGTGCTTCAATAGCTGTGGACC 1056  
QY 61 CATGATGAAGAGCAGATCTTGTATAGGCGCCACTGTAGATCAGATGAAGACTCTGA 120  
DB 1057 CATGATGAAGAGCAGATCTTGTATAGGCGCCACTGTAGATCAGATGAAGACTCTGA 1116  
QY 121 CAACAGTGCATTTATGTACAAAGATTAAATAGACAGTGTGACTCTAGATGATCTGGAGA 180  
DB 1117 CAACAGTGCATTTATGTACAAAGATTAAATAGACAGTGTGACTCTAGATGATCTGGAGA 1176  
QY 181 CTTCTTTAAGAGGTGGGTGTTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
DB 1177 CTTCTTTAAGAGGTGGGTGTTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 1236  
QY 241 CATCTACCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 300  
DB 1237 CATCTACCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 1296  
QY 301 CCCACCACTGCAAGGCTGCGGTGGATGGTTTGTATGGGAAGATTTTCAAGGAGCAA 360  
DB 1297 CCCACCACTGCAAGGCTGCGGTGGATGGTTTGTATGGGAAGATTTTCAAGGAGCAA 1356  
QY 361 ACTTAAAGTCTCCCTTGTGTCGGAAGAAAGCTTCATGAACAGTATGGGGGTGTCTGCC 420  
DB 1357 ACTTAAAGTCTCCCTTGTGTCGGAAGAAAGCTTCATGAACAGTATGGGGGTGTCTGCC 1416  
QY 421 ACCCGTAGAGGAGGAGGATGCCACCACTCCGTGAGAGTCCAGGAGGCCAGGAGG 480

DB 1417 ACCCGTAGAGGAGGAGGATGCCACCACTCCGTGAGAGTCCAGGAGGCCAGGAGG 1476  
QY 481 TCCCTGGGGAGACCCCATGGGTGCGATGGAGAGCCGTGAGAGAGATAGAGAGGCTTCCCTCC 540  
DB 1477 TCCCTGGGGAGACCCCATGGGTGCGATGGAGAGCCGTGAGAGAGATAGAGAGGCTTCCCTCC 1536  
QY 541 AAGAGGACCCCGGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
DB 1537 AAGAGGACCCCGGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1596  
QY 601 TGGAGACTGCAAGTGTCCCAATCCGGGTGTGAAAACAGAACTTGCCTGGAGAACAG 660  
DB 1597 TGGAGACTGCAAGTGTCCCAATCCGGGTGTGAAAACAGAACTTGCCTGGAGAACAG 1656  
QY 661 GTGCAACCAAGTGTAAAGCCCAAGGCTTAAGGCTTCCCGCCACCCCTTCCGGCCCC 720  
DB 1657 GTGCAACCAAGTGTAAAGCCCAAGGCTTCCCGCCACCCCTTCCGGCCCC 1716  
QY 721 GGGTGTGATCGTGGAGAGGTGGCCCTGTGATGCGGGAGAGAGTGGCTCAT 780  
DB 1717 GGGTGTGATCGTGGAGAGGTGGCCCTGTGATGCGGGAGAGAGTGGCTCAT 1776  
QY 781 GGATCGTGTGTGTCCGATGGAATTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 840  
DB 1777 GGATCGTGTGTGTCCGATGGAATTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 1836  
QY 841 CCGTGTGGCCGGGAGATGAGACCGAGAGTGGCTTGTGAGAGAAACAGAGTGGCCCTGG 900  
DB 1837 CCGTGTGGCCGGGAGATGAGACCGAGAGTGGCTTGTGAGAGAAACAGAGTGGCCCTGG 1896  
QY 901 GGGGCCCCCTGACCTTGTGAGAACAGATGGAGAGAGAGAGAGTGGAGAGC 960  
DB 1897 GGGGCCCCCTGACCTTGTGAGAACAGATGGAGAGAGAGAGAGTGGAGAGC 1956  
QY 961 TGGAAAATGATTAAGAGGAGACCGTCAAGAGCCAGAGATGGGCTTAC 1012  
DB 1957 TGGAAAATGATTAAGAGGAGACCGTCAAGAGCCAGAGATGGGCTTAC 2008

RESULT 5  
LOCUS CR608375 2314 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D1039YJ24 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR608375  
VERSION CR608375.1 GI:50489182  
KEYWORDS HTC; CNSLT\_CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2314)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2314)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source location/Qualifiers  
1..2314  
/organism="Homo sapiens"





Oy		8411	CCGTGATGGCCCGGGCAGTGAACGACCGAGTGCGCTTTGGTGGAGAAACAAGGTGGCCCTMG	900
Dd		1782	CCGTGATGGCCCGGGCAGTGAACGACCGAGTGCGCTTTGGTGGAGAAACAAGGTGGCCCTMG	1841
Oy		901	GGGGCCCCCTGGACCTTTGATGAAACAGATGGAGAAAAGAGAGAGACGTGGAGACC	960
Dd		1842	GGGGCCCCCTGGACCTTTGATGAAACAGATGGAGAAAAGAGAGAGACGTGGAGACC	1901
Oy		961	TGGAAAAATTGATAAAGCGAGCACCGTCAGAGAGCGCAGAATCGGCCCTTAC	1012
Dd		1902	TGGAAAAATTGATAAAGCGAGCACCGTCAGAGAGCGCAGAATCGGCCCTTAC	1953
	RESULT 8			
	CR610888			
	LOCUS			
	DEFINITION	CR610888	2126 bp	mRNA linear HTC 21-JUL-2004
				(human).
	ACCESSION	CR610888		
	VERSION	CR610888.1	GI:50491695	
	KEYWORDS	HTC; CNSLT cDNA.		
	SOURCE	Homo sapiens (human)		
	ORGANISM	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	REFERENCE			
	AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
	TITLE	Full-length cDNA libraries and normalization		
	JOURNAL	Unpublished		
	REMARK	Contact : Feng liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paradise Avenue 2 (bases 1 to 2126) Genoscope. Direct Submission Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefegenoscope.cns.fr - Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen. Location/Qualifiers 1..2126 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSODP037YI07" /tissue.type="Fetal brain" /plasmid="pCMVSPORT_6"		
	FEATURES			
	SOURCE			
	ORIGIN			
	Query Match	98.4%	Score 996;	DB 3; Length 2126;
	Best Local Similarity	98.8%;	Pred. No. 7.8e-237;	
	Matches 1000;	Conservative	0; Mismatches 12;	Indels 0; Gaps 0;
Oy		1	AGGACGCGGTGATGATGGACGCGCTGGAGAGCGAGTGGCTTCAATAAGCCTGGTGACC	60
Dd		942	AGGAGGACGCGGTGATGATGGCGCTGGAGAGCGAGTGGCTTCAATAAGCCTGGTGACC	1001
Oy		61	CATGATGAAGGACACAGATCTTGATCTAGGCCACCTTAGATCCAGATGAAGACTGGA	120
Dd		1002	CATGATGAAGGACACAGATCTTGATCTAGGCCACCTTAGATCCAGATGAAGACTGGA	1061
Oy		121	CAACAGTGCATTTATGTACCAAGATTAAATGACAGTGTGACTGTAGATGCTGGCAGA	180
Dd		1062	CAACAGTGCATTTATGTACCAAGATTAAATGACAGTGTGACTGTAGATGCTGGCAGA	1121
Oy		181	CTTCTTTAAGCAGTGTGGGGTTTGTTAAGATGAACAAGAACCTGGGCAACCATGATCCA	240
Dd		1122	CTTCTTTAAGCAGTGTGGGGTTTGTTAAGATGAACAAGAACCTGGGCAACCATGATCCA	1181

QY	241	CATCTACTGGACAAAGAAACAGGAAAGCCCAAGAGCGCATGCCACAGTCTCTTATGAGA	300
Db	1182	CATCTACTGGACAAAGAAACAGGAAAGCCCAAGAGCGCATGCCACAGTCTCTTATGAGA	1241
QY	301	CCCAACCACTCCCAAGGCTGCGGTGATATGTTGATGGAAATATTTTCAAGGAGCAA	360
Db	1242	CCCAACCACTCCCAAGGCTGCGGTGATATGTTGATGGAAATATTTTCAAGGAGCAA	1301
QY	361	ACTTAAAGTCTCCCTTGCTGGAAAGAGCTTCCATGACATATGCGAGGAGTGTCTGCC	420
Db	1302	ACTTAAAGTCTCCCTTGCTGGAAAGAGCTTCCATGACATATGCGAGGAGTGTCTGCC	1361
QY	421	ACCCCGTAGGGGACAGAGGCATGCCACCACTCCGTGAGAGTCCAGAGGCCCAAGAGG	480
Db	1362	ACCCCGTAGGGGACAGAGGCATGCCACCACTCCGTGAGAGTCCAGAGGCCCAAGAGG	1421
QY	481	TCTGTGGGGGACCCATGAGGTGCGATGAGAGGCGCGTGAGAGATAGAGAGGCTTCCCTCC	540
Db	1422	TCTGTGGGGGACCCATGAGGTGCGATGAGAGGCGCGTGAGAGATAGAGAGGCTTCCCTCC	1481
QY	541	AAGAGGACCCCGGGGGTTCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC	600
Db	1482	AAGAGGACCCCGGGGGTTCCGAGGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC	1541
QY	601	TGAGAGCTGGCAGTGTCCCAATCCGGGGTGTGAGAAACAGAACTTCGCTGGAAACAGA	660
Db	1542	TGAGAGCTGGCAGTGTCCCAATCCGGGGTGTGAGAAACAGAACTTCGCTGGAAACAGA	1601
QY	661	GTGCACACAGTGTAAAGGCCCAAGACCTGAAAGGCTTCTCCCGCACCTTTCCGCCCC	720
Db	1602	GTGCACACAGTGTAAAGGCCCAAGACCTGAAAGGCTTCTCTCCCGCACCTTTCCGCCCC	1661
QY	721	GCGTGTGATGTGTGGCAGAGGTGGCCCTGTGTCATGCGGGGAGGAAAGAGGTGGCTCAT	780
Db	1662	GCGTGTGATGTGTGGCAGAGGTGGCCCTGTGTCATGCGGGGAGGAAAGAGGTGGCTCAT	1721
QY	781	GGATCGTGTGTGTCCCGGTGGAAATGTTCAAGGTGGCGGTGTGAGACAGAGGTGGCTT	840
Db	1722	GGATCGTGTGTGTCCCGGTGGAAATGTTCAAGGTGGCGGTGTGAGACAGAGGTGGCTT	1781
QY	841	CCGTGTGTGCGCGGGGACATGACCCGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCCCTGG	900
Db	1782	CCGTGTGTGCGCGGGGACATGACCCGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCCCTGG	1841
QY	901	GGGGCCCCCTTGACCTTTGATGAAACATGTGGAGAAAGAAAGAGAGAGCGTGGAGAAC	960
Db	1842	GGGGCCCCCTTGACCTTTGATGAAACATGTGGAGAAAGAAAGAGAGAGCGTGGAGAAC	1901
QY	961	TGGAAAAATGGATTAAGGCGAGCAACCGTCAAGAGCGGAGAGATGCGGCCCTTAC	1012
Db	1902	TGGAAAAATGGATTAAGGCGAGCAACCGTCAAGAGCGGAGAGATGCGGCCCTTAC	1953
RESULT 9			
CR625247			
LOCUS	full-length cDNA clone CSDB0051F11 of Placenta of Homo sapiens		
DEFINITION	full-length cDNA clone CSDB0051F11 of Placenta of Homo sapiens (human).		
ACCESSION	CR625247		
VERSION	CR625247.1 GI:50506054		
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Emmalyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1. (bases 1 to 2159)		
AUTHORS	Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished		
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue		

REFERENCE 2 (bases 1 to 2159)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 FEATURES  
 source 1. 2159  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D05YF11"  
 /tissue\_type="Placenta"  
 /plasmid="pCMVSPORT\_6"

ORIGIN

Query Match 98.4%; Score 996; DB 3; Length 2159;  
 Best Local Similarity 98.4%; Pred. No. 7.8e-237;  
 Matches 1000; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

1 AGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 60  
 |||||  
 999 AGGAGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 1058  
 |||||  
 61 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120  
 |||||  
 1059 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 1118  
 |||||  
 121 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATATCTGGCAGA 180  
 |||||  
 1119 CAACAGTGCATTTATGTACAGAGATTAAATGACAGTGTGACTCTAGATATCTGGCAGA 1178  
 |||||  
 181 CTTCTTTAAGCAGTGTGGGTTTGAATGATGAACAAGAACTGGGGAACCCATGATCCA 240  
 |||||  
 1179 CTTCTTTAAGCAGTGTGGGTTTGAATGATGAACAAGAACTGGGGAACCCATGATCCA 1238  
 |||||  
 241 CATCTACCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 300  
 |||||  
 1239 CATCTACCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAGA 1298  
 |||||  
 301 CCCACCACTGCCAAGGCTGCGTGAATGTGATGGAAAGATTTTCAAGGAGCAA 360  
 |||||  
 1299 CCCACCACTGCCAAGGCTGCGTGAATGTGATGGAAAGATTTTCAAGGAGCAA 1358  
 |||||  
 361 ACTTAAAGTCTCTTGTCTGGAAGAAAGCTTCCATGAACATGTGGGGGTGCTTGGC 420  
 |||||  
 1359 ACTTAAAGTCTCTTGTCTGGAAGAAAGCTTCCATGAACATGTGGGGGTGCTTGGC 1418  
 |||||  
 421 ACCCGTGAAGGAGGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAG 480  
 |||||  
 1419 ACCCGTGAAGGAGGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAG 1478  
 |||||  
 481 TCCTGGGAGGACCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 540  
 |||||  
 1479 TCCTGGGAGGACCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 1538  
 |||||  
 541 AAGAGACCCCGGGGTTTCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGAC 600  
 |||||  
 1539 AAGAGACCCCGGGGTTTCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGAC 1598  
 |||||  
 601 TGGAGACTGGCAGTGTCCCAATCCGGGTTTGAAGAACAGAACTTGGCTGAGAAACAGA 660  
 |||||  
 1599 TGGAGACTGGCAGTGTCCCAATCCGGGTTTGAAGAACAGAACTTGGCTGAGAAACAGA 1658  
 |||||  
 661 GTGCAACCAAGTGTAGAGGCCCAAGGCTTCCCTCCGACCCCTTTCCGCCCCC 720  
 |||||  
 1559 GTGCAACCAAGTGTAGAGGCCCAAGGCTTCCCTCCGACCCCTTTCCGCCCCC 1718  
 |||||

ORIGIN

Query Match 88.1%; Score 892; DB 3; Length 2236;  
 Best Local Similarity 98.7%; Pred. No. 6.2e-211;  
 Matches 896; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

1 AGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 60  
 |||||  
 972 AGGAGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 1031  
 |||||  
 61 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120  
 |||||  
 1032 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 1091  
 |||||

REFERENCE 10  
 CR604639 2236 bp mRNA linear HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS0D013YN08 of T cells (Jurkat cell line)  
 DEFINITION Cot 10-normalized of Homo sapiens (human).  
 VERSION CR604639.1 GI:50485446  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2236)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue  
 2 (bases 1 to 2236)

REFERENCE 2 (bases 1 to 2236)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 FEATURES  
 source 1. 2236  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D013YN08"  
 /tissue\_type="T cells (Jurkat cell line) Cot 10-normalized"  
 /plasmid="pCMVSPORT\_6"

ORIGIN

Query Match 88.1%; Score 892; DB 3; Length 2236;  
 Best Local Similarity 98.7%; Pred. No. 6.2e-211;  
 Matches 896; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

1 AGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 60  
 |||||  
 972 AGGAGGACGGGTGGATGGGACGCGCTGGAGAGCGAGTGGCTCAATAGCTGTGGACC 1031  
 |||||  
 61 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120  
 |||||  
 1032 CATGATGAGAGACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 1091  
 |||||

OY		121	CAACAGTGCATTTTAATGTACAAGAAATTAAATGACAGTGGACTCTAATGATATCTGGCAGA	180
Db		1092	CACACGTGCATAATTATATGTACAAAGATTAAATATACAGTGTGACTCTAATGATATCTGGCAGA	1151
OY		181	CTTCTTTAAGCAGTGTGGGGTGTTTAAGATGAAACAGAAACATCGGSCAACCCATATCCA	240
Db		1152	CTTCTTTAAGCAGTGTGGGGTGTTTAAGATGAAACAGAAACATCGGSCAACCCATATCCA	1211
OY		241	CATCTAATCTGGACAAAGAAAACAGAAAGGCCCAAAGCCGATGCCACAGTGTCTTATGAAAGA	300
Db		1212	CATCTAATCTGGACAAAGAAAACAGAAAGGCCCAAAGCCGATGCCACAGTGTCTTATGAAAGA	1271
OY		301	CCCACCACACTGTCAGAGGCGCCGTGGAAATGTTTTGATGAGAAAGATTTTCAAGGGAGACA	360
Db		1272	CCCACCACACTGTCAGAGGCGCCGTGGAAATGTTTTGATGAGAAAGATTTTCAAGGGAGACA	1331
OY		361	ACTTAAAGTCTTCCTCTGTCTCGAAGAAAGCCTCCAAATGAAACAGTATGCGGGGTGTCTGCC	420
Db		1332	ACTTAAAGTCTTCCTCTGTCTCGAAGAAAGCCTCCAAATGAAACAGTATGCGGGGTGTCTGCC	1391
OY		421	ACCCCGTGAAGGACAGAGGCAATGCCACACCACTCCGTGAGGTCCAGAGGCCCAAGAGG	480
Db		1392	ACCCCGTGAAGGACAGAGGCAATGCCACACCACTCCGTGAGGTCCAGAGGCCCAAGAGG	1451
OY		481	TCTCTGGGGGACCCATGGGTCGCAATGGAGAGGCGGTGAGAGGATAGAGGAGGCTTCCCATC	540
Db		1452	TCTCTGGGGGACCCATGGGTCGCAATGGAGAGGCGGTGAGAGGATAGAGGAGGCTTCCCATC	1511
OY		541	AAGAGGACCCCGGGGTTCCCGAAGGAAACCCCTCTGAGAGAGGAAACGTCCAGACCGAGC	600
Db		1512	AAGAGGACCCCGGGGTTCCCGAAGGAAACCCCTCTGAGAGAGGAAACGTCCAGACCGAGC	1571
OY		601	TGGAAATCTGGCAGTGTCCCAATCCGGGTGTGTGAAACCAAGAACTTTCGCTGTGAAACAGA	660
Db		1572	TGGAAATCTGGCAGTGTCCCAATCCGGGTGTGTGAAACCAAGAACTTTCGCTGTGAAACAGA	1631
OY		661	GTCGAACAGATGTTAAGGCCCAACAAACCTGAAAGGCTTCCGCCGACACCCCTTTCGSCCCCC	720
Db		1632	GTCGAACAGATGTTAAGGCCCAACAAACCTGAAAGGCTTCCGCCGACACCCCTTTCGSCCCCC	1691
OY		721	GGGTGTGATCTGTGCAGAGGTGTGCCCTTGTGTGCATGCGGGAGAGAAAGTGTGCCCTCAT	780
Db		1692	GGGTGTGATCTGTGCAGAGGTGTGCCCTTGTGTGCATGCGGGAGAGAAAGTGTGCCCTCAT	1751
OY		781	GGATCTGTGTGTCTCCCGGTGAAATGTTCAGAGGTGTGCCGTGTGTGAGACAGAGTGTGCTT	840
Db		1752	GGATCTGTGTGTCTCCCGGTGAAATGTTCAGAGGTGTGCCGTGTGTGAGACAGAGTGTGCTT	1811
OY		841	CCGTGTGTGACCGGGGACATGAAACCAAGATGTGCTTTGTGTGAGAAACAGAGGTGTGCCCTG	900
Db		1812	CCGTGTGTGACCGGGGACATGAAACCAAGATGTGCTTTGTGTGAGAAACAGAGGTGTGCCCTG	1871
OY		901	GGGGGCCCC G08	
Db		1872	GGGGGCCCC 1879	
<b>RESULT 11</b>				
AK014366				
LOCUS				
DEFINITION				
AK014366 1379 bp mRNA linear HTC 03-APR-2004				
Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched				
library, clone:3300002D11 product:Ewing sarcoma homolog, full				
insert sequence.				
ACCESSION				
AK014366				
VERSION				
AK014366.1 GI:12852167				
KEYWORDS				
HTC; CAP trapper.				
SOURCE				
Mus musculus (house mouse)				
ORGANISM				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.				
REFERENCE				
1 Carninci,P. and Hayashizaki,Y.				
AUTHORS				

TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	992792925
PubMed	10349636
REFERENCE	
AUTHORS	2
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)
PubMed	10499374
REFERENCE	3
AUTHORS	11042159
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kibunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sataguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
JOURNAL	RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)
PubMed	20530913
REFERENCE	4
AUTHORS	11076861
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL	Functional annotation of a full-length mouse cDNA collection
MEDLINE	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 1379)
REFERENCE	
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirao, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kaubawa, T., Kato, H., Kawai, T., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyata, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
MEDLINE	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
REFERENCE	Please visit our web site ( <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> ) for further details.
AUTHORS	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGACAGATCCAGAGACTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAATTAATGCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI. Cloning sites, 5' end: XhoI, 3' end: SctI. Host: SOLR.
COMMENT	1. Location/Qualifiers
FEATURES	
source	1..1379



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:3300002D11"
/db_xref="taxon:10090"
/clone_xref="3300002D11"
/feature_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="17 days embryo"
<1..1004
note="unnamed protein product; Ewing sarcoma homolog
(MGD|MG1:99960, GB|NM_007968, evidence: BLASTN, 99%,
match=2172)
putative"
/codon_start=3
/protein_id="BAB29301.1"
/db_xref="GI:12852168"
/translation="LQSGAGRGGRGPKRGGPMDEGPDLDLGLPIDPEDSDNSAIYYQG
LNDNVTLDLADLPKOCGVKMKRGOPIHITYLDEKGRKPKDADTVSEYEDPTAKA
AYEFDGKDFQSGKLYSLARKKPRPMASNRGMPPRBRGRMPPLRGGRGPGPGPGP
MGRMGGRGDRGSPPRGPRGSRNBSGNSVDRAGDWCNPPGCGNDFARTCN
QCKAPRBEGLPPPPPPGDRGRGGRGGLMDRGCGMFRGGRGGDRGF
RGRGMDRGFGGRRGPGPGPIEMQWGRRGGRGGRGKMDKSHRERRDRBY"
1351..1356
note="putative"
1379
note="putative"

polysite
1379
note="putative"

ORIGIN
Query Match      87.2%; Score 882.6; DB 3; Length 1379;
Best Local Similarity 92.6%; Pred. No. 1,2e-208;
Matches 927; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 12 GAATGGGACGCGTGGAGAGGAGTGGCTTCAATAAGCTGTGGAGCCATGATGAG 71
DB 1 GACTGGGACGCGTGGAGAGGAGTGGCTTCAATAAGCTGTGGAGCCATGATGAG 60

QY 72 GACCAAGTCTTGATCTAGGCCCACTGTGATGATCCAGATGAAGACTCGAACAAGTGC 131
DB 61 GACCAAGTCTTGATCTAGGCCCTTCTTATGATCCAGATGAAGACTCGAACAAGTGC 120

QY 132 TTATATGACAAGATTAATGACAGTGTGACTGATGATGATCTGGCAGACTTCTTAAAG 191
DB 121 TTATATGACAAGATTAATGACAGTGTGACTGATGATGATCTGGCAGACTTCTTAAAG 180

QY 192 AGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCAATCTTACCTGG 251
DB 181 AGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCAATCTTACCTGG 240

QY 252 ACAAGGAAACAGGAAACCCCAAGGCGATGCCAAGATGCTCTATGAAGACCCACCTG 311
DB 241 ATAAAGGAAACAGGAAACCCCAAGGCGATGCCAAGATGCTCTATGAAGATCCACCACTG 300

QY 312 CCAAGGCTGCGGTGAATGTTTATGAGGAAAGATTTTCAAGGAGGACAACTTAAAGTCT 371
DB 301 CCAAGGCTGCGGTGAATGTTTATGAGGAAAGATTTTCAAGGAGGACAACTTAAAGTCT 360

QY 372 CCCTTGTGCGAAGAGCTTCAATGAACAGTATGCGGGGTGATCTGCCACCCCGTAGG 431
DB 361 CTCTTGTGCGAAGAGCTTCAATGAACAGTATGCGGGGTGATCTGCCACCCCGTAGG 420

QY 432 GCAGAGGATGCCCAACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGGTCTTGGGGAC 491
DB 421 GCAGAGGATGCCCAACCACTCCGTGAGAGTCTTGGGGGTGAGAGGTCTTGGGGAC 480

QY 492 CCATGGGTGCGATGAGAGGCGGTGAGAGGATGAGAGGCTTCCCTCCCAAGAGAGCC 551
DB 481 CCATGGGTGCGATGAGAGGCGGTGAGAGGATGAGAGGCTTCCCTCCCAAGAGAGCC 540

QY 552 GGGGTTCCCGAGGAAACCCCTCTGAGAGAGGAAAGTTCAGACACCGAGCTGAGAGTGGC 611
DB 541 GAGGCTCCAGAGGAAACCCCTCTGAGAGAGGAAAGTTCAGACACCGAGCTGAGAGTGGC 600

```

```

QY 612 AGTGTCCCAATCCGGGTGTGGAACACAGAACTTGCGTGGAGAACAGATGCACCACT 671
DB 601 AGTGTCCCAATCCGGGTGTGGAACACAGAACTTGCGTGGAGAACAGATGCACCACT 660

QY 672 GTAAAGGCCCAAGAGCTGGAAGCTTCTCCGACACCTTTTCCGCCCGGGGTGTGATC 731
DB 661 GTAAAGGCCCAAGAGCTTCTCCGACACCTTTTCCGCCCGGGGTGTGATC 720

QY 732 GTGCGAGAGTGTGCTGTGTCAGATGCGGAGGAGAAAGATGTGCTCAATGATCGTGTG 791
DB 721 GTGCGAGAGTGTGCTGTGTCAGATGCGGAGGAGAAAGATGTGCTCAATGATCGTGTG 780

QY 792 GTCCCGGTGAATGTTCAAGAGTGTGCGCGTGTGAGACAGAGTGTGCTTCCGTGTGCC 851
DB 781 GTCCCGGTGAATGTTCAAGAGTGTGCGCGTGTGAGACAGAGTGTGCTTCCGTGTGCC 840

QY 852 GGGGCAATGACCGAGGTGCTTTGTGTGAGAGAAACAGAGTGTGCTTGGGGGGCCCTG 911
DB 841 GTGGAATGACCGAGGTGCTTTGTGTGAGAGAAACAGAGTGTGCTTGGGGGGCCCTG 900

QY 912 GACCTTTGATGAAACAGATGGAGAGAGAAAGAGAGAGAGTGTGAGGACCTGGAAAAATG 971
DB 901 GACCTTTGATGAAACAGATGGAGAGAGAAAGAGAGAGAGTGTGAGGACCTGGAAAAATG 960

QY 972 ATAAAGCGAGACCGCTCAGAGCGAGAGATGCGGCTTAC 1012
DB 961 ATAAAGCGAGACCGCTCAGAGCGAGAGATGCGGCTTAC 1001

RESULT 12
AK019460      2107 bp      mRNA      linear      HTC 03-Apr-2004
LOCUS
DEFINITION
Mus musculus 18 days pregnant adult female placenta and extra
embryonic tissue cDNA, RIKEN full-length enriched library,
clone:3830417B11 product:Ewing sarcoma homolog, full insert
sequence.
ACCESSION
AK019460
VERSION
AK019460.1 GI:12859676
KEYWORDS
HTC, CAP trapper (house mouse)
SOURCE
Mus musculus
ORGANISM
Mus musculus
REFERENCE
1. Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
REFERENCE
2. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
REFERENCE
3. Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, I., Matsuura, S., Kawai, D.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
REFERENCE
4. The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

```

**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL REFERENCE** Nature 409, 685-690 (2001)  
**AUTHORS** 5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**TITLE** Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL REFERENCES** Nature 420, 563-573 (2002)  
**AUTHORS** 6 (bases 1 to 2107)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Camici, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirotsu, T., Horii, F., Imotani, K., Ishii, Y., Itch, M., Izawa, M., Kaekawa, T., Kato, H., Kawai, Y., Kojima, Y., Komori, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT** Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGAGAAGCATCCAGAGCTCTTTTCTTTTCTTTTA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adtercent of sequence [5' GAGAGGAGATTCACGATTAAATTATATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmidscript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

**FEATURES**  
**SOURCE** Location/Qualifiers  
1..2107  
    `/organism="Mus musculus"`  
    `/mol_type="mRNA"`  
    `/strain="C57BL/6J"`  
    `/db_xref="PANTOM.DB:3830417B11"`  
    `/db_xref="taxon:10090"`  
    `/clone="3830417B11"`  
    `/sex="female"`  
    `/tissue_type="Placenta and extra embryonic tissue"`  
    `/clone_lib="RIKEN full-length enriched mouse cDNA library"`  
    `/dev_stage="18 days pregnant adult"`  
17..1984  
    `/note="Unnamed protein product; Ewing sarcoma homolog [MGI:99960, GB|NM_007968, evidence: BLASTN, 99%, match=2172]`  
    `putative"`  
    `/codon_start=1`  
    `/protein_id="BAB31732.1"`  
    `/db_xref="GI:1285677"`  
    `/translation="MASTDYSTYQAAAQGYSAVTAOPTGVACTOAYGOOSYGTYT  
GOPDYSIAQTAVTGQATVGSQATVATSYSPPTGPAPASQPVVGSGYGTADSTT  
ATVTTSVAQAOTAVGPAVPYTGQOPPTATPRPDGKRAETSSPOPSOONYNO  
PSLGGSNVSYPDVPGSPMPQTPAAPSPPSVSSPSOPNSKPSOONNYGDS  
SYGOOSISYGOOSITGQQPPTSYPQTGSYSQAPSYCOOSSSYQOQSFRDDHPSNG  
VYGDSEGFSGPGENRSLSGEDNRGRGDEDRGMRSRGRRGGGGLGAGERGNPK  
PGWGEDEPDIDLPIIDPDEDSDNSAIYVGILNDNTLDLPDFKQGVVNKKRT`

Query Match	86.9%	Score 879.2	DB 3	Length 2107
Best Local Similarity	91.7%	Pred. No. 9.4e-208		
Matches 928	Conservative 0	Mismatches 84	Indels 0	Gaps 0
QY 1	AGGACGGGCTGGAATGGGCAAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGGTGGAC	60		
DB 970	AGGAGGACGCGGCTGGAATGGGCAAGCGGCTGGAGAGCGGCTTCAATAGCTGGTGGAC	1029		
QY 61	CATGATGAGAGGACCAAGATCTTGATCTTAGGCGCCACCTGATGATCCAGATGAAGACTCTGA	120		
DB 1030	CATGATGAGAGGACCAAGATCTTGATCTTAGGCGCTTCTATGATCCAGATGAAGACTCTGA	1089		
QY 121	CAACAGTGCATTTATGTACAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	180		
DB 1090	CAACAGTGCATTTATGTACAAGATTAAATGACAGTGTGACTCTGATGATCTGGCAGA	1149		
QY 181	CTTCTTTAAGAGTGTGGGCTTTTAAGATGAAACAAGAACTGGGCAACCCATGATCCA	240		
DB 1150	CTTCTTTAAGAGTGTGGGCTTTTAAGATGAAACAAGAACTGGGCAACCCATGATCCA	1209		
QY 241	CATCTACCTTGACAAGAGAAACAGGAAAGCCCAAGGGGATGCGACAGTGTCTTATGAGA	300		
DB 1210	TATCTACTCTGATTAAGAGACAGGAAAGCTTAAGGGGACGCCACAGTGTCTTATGAGA	1289		
QY 301	CCCAACCACTGCCAAGGCTGCGCTGGAATGGTTGATGGAGAAAGATTTTCAAGGAGCAA	360		
DB 1270	TCCAACCACTGCCAAGGCTGCGCTGGAATGGTTGATGGAGAAAGATTTTCAAGGAGCAA	1329		
QY 361	ACTTAAAGTCTCTCTTGTCTGGAAAGAGCTTCCATGAACATATGCGGGGTGTCTTGCC	420		
DB 1330	ACTTAAAGTCTCTCTTGTCTGGAAAGAGCTTCCATGAACAGCATGTGGGGAGGCAATGCC	1389		
QY 421	ACCCCGTAGGGGACAGAGGATGCGCACCACTCCGAGAGGTCCAGAGGCGCCAGAGG	480		
DB 1390	ACCTCGTAGGGGACAGGGGAGATGCCACCACTTCTGTGAGGTCTGTGGCCAGAGG	1449		
QY 481	TCCTGGGGGACCCATGGGTGCGATGGAGGCGCTGAGAGAGATGAGAGGCTTCCCTCC	540		
DB 1450	CCCTGGAGGACCCATGGGTGCGATGGAGGCGCTGAGAGAGATGAGAGGCTTCCCTCC	1509		
QY 541	AAGAGAACCCCGGGGCTTCCGAGGGAAACCTCTGTGAGAGGAAAGTCCAAGCCGAGC	600		
DB 1510	AAGAGGGGCGCCGAGGCTTCCAGAGGAAACCTCTGTGAGAGGAAAGTCCAAGCCGAGC	1569		
QY 601	TGGAGACTGGCAGTGTCCCAATCCGGGCTTGAGAAACAAGACTTGTGCGCTGGAGACAG	660		
DB 1570	TGGAGACTGGCAGTGTCCCAATCCGGGCTTGAGAAACAAGACTTGTGAGAAACAG	1629		
QY 661	GTGCAACCAAGTGAAGGCCCAAGACTGGAAGGCTTCTCCCGCACCTTTCCGCGCCC	720		
DB 1630	ATGCAACCAAGTGAAGGCCCTTGAAGGCCGAGGCTTCTCCCGCACCTTTCCACCTCC	1689		
QY 721	GAGGTGTGATCTGTGACAGAGTGGCCCTGTGTGATGCGGGAGAGAAAGAGTGGCTCAT	780		
DB 1690	GAGGTGTGATCTGTGACAGAGTGGCCCTGTGTGATGCGGGAGAGAAAGAGTGGCTCAT	1749		
QY 781	GGAATCGTGTGTCTCCCGTGAAGATGTTCAAGAGTGGCGCTGTGTGAGACAGAGTGGCTT	840		
DB 1750	GGAATCGTGTGTCTCTGAGAGAAATGTTCAAGAGTGGCGAGAGTGTGAGACAGAGAGGCTT	1809		
QY 841	CCGATGTGTGCGGGGACATGACCGAGGTGCTTTGTGTGAGAGAAAGAGTGTGCGCTTGG	900		
DB 1810	CCGATGTGTGCGCTGTGAGATGAGACCGAGGTGGCTTTGTGTGAGAGAAAGAGTGTGCG	1869		
QY 901	GAGGCGCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGCTGTGAGAGCC	960		

Db	1870	GGGGGCTCCTGGACCTTTAATGGAACAGANTGGAGAGAAAGAGCGGACGCGAGGACC	1992
Qy	961	TGGAAAAATGATTAAGCGCAGACCGCTCAGAGCGCGAGAGATCGGCCCTTAC	1012
Db	1930	TGGGAAAAATGATTAAGCGCAGACCGCTCAGAGAAAGAGCGGACCGGCCCTTAC	1981
RESULT 13			
LOCUS	AK034755	2373 bp	mRNA
DEFINITION	Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430031m11		
ACCESSION	AK034755.1	GI:26084174	
VERSION	AK034755.1	GI:26084174	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	High. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subrestriction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	2049374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Atzawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Teshiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format		
JOURNAL	sequencing pipeline with 384 multicapillary sequencer		
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5		
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
REFERENCE	6 (bases 1 to 2373)		
AUTHORS	Adachi, J., Atzawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kojima, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahata, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-Jul-2001) Yoshinide Hayashizaki, The Institute of		

COMMENT	FEATURES	ORIGIN
Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216, Fax: 81-45-503-9216)	source 1. .2373 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM_DB:9430031M11" /db_xref="taxon:10090" /clone="9430031M11" /tissue_type="embryonic body between diaphragm region and neck" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 43. .2011 /note="Ewing sarcoma homolog (MGD) MGI:99960, GB NM_007968, evidence: BLASTN, 99%, match=2172) putative" 2358. .2363 /note="putative" 2373 /note="putative"	Query Match 86.6%; Score 876; DB 3; Length 2373; Best Local Similarity 91.5%; Pred. No. 6.1e-207; Matches 926; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
	misc_feature polyA_signal polyA_site	
	1 AGGACGCGGATGGAATGGGACGCGCTGGAGACGAGTGGCTTCATTAAGCCTGGTGACC 60     997 AGGAGACGCGGTGAGCTGGGCGCTGGAGACGAGTGGCTTCATTAAGCCTGGTGACC 10566 	
	61 CATGGATGAAGACCAAGATCTTGATCTTAGGCCCACTGTAGATCCAGTGAAGACTCTGA 120     1057 CATGGATGAAGACCAAGATCTTGATCTTAGGCCCTTCTTAGATCCCGATGAAGACTCTGA 1116 	
	121 CAACAGGCAATTTATGTACAAGATTTAAAGAAGTGTACTAGATGTATCTGGCAGA 180     1117 CAACAGTCAATTTATGTACAAGATTTAAAGAAGTGTACTAGATGTATCTGGCAGA 1176 	
	181 CTTCTTTAAGAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCA 240     1177 CTTCTTTAAGAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCA 1236 	
	241 CATCTACCTGGACAAGAAACAGAAAAGCCAAAGGCGATGCCAAGTGTCTTAGAAGA 300     1237 TATCTACTGTGATTAAGAGACAGAAAAGCCTAAAGGGGACCCCAAGTGTCTTAGAAGA 1296 	
	301 CCCACCCACTGCCAAGGTCGCGGTGGAATGTTTGAATGGAAAGATTTTCAAGGAGCA 360     1297 TCCACCAACTGCCAAGGTCGCGGTGGAATGTTTGAATGGAAAGATTTTCAAGGAGCA 1356 	
	361 ACTTAAAGTCCCTGTGCTGGAGAAAGCCCTCAATGAACAGTATGGCGGGTGTCTGCC 420     1357 ACTTAAAGTCTCTTGTCCCAAGAAAGCTTCAATGAACAGTATGGCGGGAGGATCTCC 1416 	
	421 ACCCGTGAAGGAGGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAG 480     1417 ACCTGTGAAGGAGGAGGAGGATGCCACCACTCCGTGAGAGTCTGTGTGGCCAGAGAG 1476 	
	481 TCTTGGGGGACCAATGGGTGCATGGAGAGCCGTGAGAGAGATGAAGAGGCTTCCCTCC 540 	

Db	1477	CCCTGGAGGAGCCCAATGGGCTCCGATGGAGGCGCTGTGAGAGACAGAGGGGGCTTCCCTCC	1536
Qy	541	AAAGAGACCCCGGGGCTTCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGCACGAGC	600
Db	1537	AAGAGGGGCCCGAGGCTCCAGAGGAAACCCCTCTGAGAGAGGAATGTCCAGCACGAGC	1598
Qy	601	TGGAGACTGGCAGGTCTCCCAATCAAGGGCTGTGGAAACAGAACTTCGTTGGAGAACGA	660
Db	1557	TGGAGACTGGCAGGTCTCCCAATCAAGGGCTGTGGAAACAGAACTTCGTTGGAGAACGA	1656
Qy	661	GTGCAACCAAGTGTAAAGCCGCCAAGACCTGAAGGCTTCTCCCGCACCCCTTCCGACCCC	720
Db	1657	ATGCAACCAAGTGTAAAGCCGCCAAGACCCGAGGGCTTCTCCCGCACCCCTTCCACCTCC	1716
Qy	721	GGGTGTGATCTGTGCAAGAGTGCGCTGTGTGTCATGCGGGAGAGAAAGGTGGCTCAT	780
Db	1717	GGGTGTGATCTGTGCAAGAGTGCGCTGTGTGTCATGCGGGAGAGAAAGAGGAGCTCAT	1776
Qy	781	GGATCGTGTGTGTCCCCGTGGAATTTTCAGAGGTGGCCGTGTGAGACAGAGGTGGCTT	840
Db	1777	GGATCGTGTGTGTCTCTTGAGGAATTTTCAGAGGTGGCAAGGTGGACAGAGGAGGCTT	1836
Qy	841	CCGTGTGTGACCGGGGCATGAGCCGAGGTGGCTTTGTGTGAGAGAAAGAGGTTGGCCCTGG	900
Db	1837	CCGAGGTGGCCGTGGAATGTACCGAGGTGGCTTTGTGTGAGAGAAAGAGGTTGGCTTGG	1896
Qy	901	GGGGCCCCCTGACCTTTGATGGAACAGATGGAGGAAAGAAAGAGAGACGTGAGGAGCC	960
Db	1897	GGGGGCTCTCTGGAACCTTTAATGGAACAGATGGAGGAAAGAAAGGCGGACCGTGGAGGACC	1956
Qy	961	TGGAAAAATGATTAAGCGAGACACCGTCAAGAGCGCAGAGATCGGCGCTTAC	1012
Db	1957	TGGAAAAATGATTAAGCGAGACACCGTCAAGAGCGCAGAGACCGGCGCTTAC	2008

RESULT 14  
BX387336/c

LOCUS	1020 bp	mRNA	linear	EST 27-APR-2004
DEFINITION	BX387336 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens			
ACCESSION	cDNA clone CS0DD004YE17 3-PRIME, mRNA sequence.			
VERSION	BX387336			
KEYWORDS	BX387336.1 GI:30455516			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 1020)			
JOURNAL	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.			
COMMENT	Full-length cDNA libraries and normalization Unpublished (2001)			
	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	2 rue Gaston Cremieux, CP 5706 - 91057 Evry cedex - FRANCE			
	Email: <a href="mailto:seqlife@genoscope.cns.fr">seqlife@genoscope.cns.fr</a> , Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a>			
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primed end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10589.f			
	For more information about this cluster, see <a href="http://www.genoscope.cns.fr/cdna?b=CS6AD001ZD03NP1&amp;c=10589.f">http://www.genoscope.cns.fr/cdna?b=CS6AD001ZD03NP1&amp;c=10589.f</a> .			

```

FEATURES
Source      Location/Qualifiers
1. 1020
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD0041YB7"
/clone_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-clis0 (dT).

```

[illegible]

RESULT_15			
AK049743			
LOCUS			
AK049743	2269 bp	mRNA	linear HTC 03-APR-2004

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone:C550046A18 product:Ewing sarcoma homology, full insert sequence.	AK049743	AK049743.1	GI:26093614	HTc, CAP trapper.	Mus musculus (house mouse)
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
1	Carninci, P. and Hayashizaki, Y.				
2	High-efficiency full-length cDNA cloning				
3	Meth. Enzymol. 303, 19-44 (1999)				
4	99279253				
5	10349636				
6	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,				
7	Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
8	Normalization and subtraction of cap-trapper-selected cDNAs to				
9	prepare full-length cDNA libraries for rapid discovery of new genes				
10	Genome Res. 10 (10), 1617-1630 (2000)				
11	20499374				
12	11042159				
13	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,				
14	Komou, H., Akiyama, J., Nishi, K., Kitenaka, T., Tashiro, H., Itoh, M.,				
15	Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,				
16	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, Y.,				
17	Fujiwara, S., Inoue, K., Ozawa, Y., Izawa, M., Ohara, E., Watabiki, M.,				
18	Yoneda, Y., Ishikawa, T., Togawa, K., Tanaka, T., Matsura, S., Kawai, J.,				
19	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
20	RIKEN Integrated Sequence Analysis (RISA) system-384-format				
21	sequencing pipeline with 384 multicapillary sequencer				
22	Genome Res. 10 (11), 1757-1771 (2000)				
23	11076861				
24	The RIKEN Genome Exploration Research Group Phase II Team and the				
25	PANTOM Consortium.				
26	Functional annotation of a full-length mouse cDNA collection				
27	Nature 409, 685-690 (2001)				
28	5				
29	The PANTOM Consortium and the RIKEN Genome Exploration Research				
30	Group Phase I & II Team.				
31	Analysis of the mouse transcriptome based on functional annotation				
32	of 60,770 full-length cDNAs				
33	Nature 420, 563-573 (2002)				
34	6 (bases 1 to 2269)				
35	Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,				
36	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,				
37	Hayashizaki, K., Hayatsu, N., Hizamoto, K., Hirooka, T., Hirozane, T.,				
38	Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,				
39	Karoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,				
40	Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M.,				
41	Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohse, N.,				
42	Okazaki, Y., Saito, R., Saitoh, H., Sasaki, C., Sakai, K., Sakazume, N.,				
43	Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shikata, T.,				
44	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,				
45	Takeda, Y., Tanaka, T., Tomaru, A., Toyo, T., Yasunishi, A.,				
46	Muramatsu, M. and Hayashizaki, Y.				
47	Direct Submission				
48	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of				
49	Physical and Chemical Research (RIKEN), Laboratory for Genome				
50	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),				
51	RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,				
52	Kanagawa 230-0045, Japan [E-mail:genome-research@riken.jp,				
53	URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222,				
54	Fax:81-45-503-9216)				
55	cDNA library was prepared and sequenced in Mouse Genome				
56	Encyclopedia Project of Genome Exploration Research Group in Riken				
57	Genomic Sciences Center and Genome Science Laboratory in Riken.				
58	Division of Experimental Animal Research in Riken contributed to				

FEATURES		source	
1. .2269		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="RANTOM.DB:C530046A18"	
		/db_xref="taxon:10090"	
		/clone="C530046A18"	
		/tissue_type="spinal cord"	
		/clone_id="RIKEN full-length enriched mouse cDNA library"	
		/dev_stage="12 days embryo"	
		48..1903	
		/note="Bwing sarcoma homolog (MGI) MGI:99960, GB NM_007968, evidence: BLASTN, 99%, match=2172)	
		putative"	
		2251..2256	
		/note="putative"	
		2269	
		/note="putative"	
ORIGIN			
Query Match		78.3%; Score 792.8; DB 3; Length 2269;	
Best Local Similarity		90.2%; Pred. No.3.1e-186;	
Matches 859; Conservative		0; Mismatches 92; Indels 1; Gaps 1;	
QY	61	CATGATGAAGACACAGATCTTGATCTTAGGCCCACTGTGTGATCCAGATGAAGACTCTGA	120
DB	950	CAGGGGAAGGGGGATTTGATCGTGAGGACATGACAGAGATCCCATGAAGACTCTGA	1009
QY	121	CAACAGTGCATTATATATACAAAGATTAATAATGACAGTGTACTGTAGATGATCTGGCAGA	180
DB	1010	CAACAGTGCATTATATATGTCAGAGATTAATAATGACATGTACTGTAGATGATCTGGCAGA	1069
QY	181	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAGAATTGGGCAACCCATGATCCA	240
DB	1070	CTTCTTTAAGCAGTGTGGGGTTGTCAGATGAACAAGAGAATTGGGCAACCCATGATCCA	1129
QY	241	CATCTACCTGGACAAAGAAACAGAAAGGCCAAAGCGATGCCAGTGTCTTATGAAGA	300
DB	1130	TATCTACCTGGATTAAGAGACAGAAAGCCTTAAAGGGGAAGCCCAAGTGTCTTATGAAGA	1189
QY	301	CCGACCACTGCCAAAGCGTCCGGTGAATGATGAGGAAAGATTTTCAAGGGACAA	360
DB	1190	TCCACCACTGCCAAGAGCTCCGTGGAATGATGAGGAAAGATTTTCAAGGACAA	1249
QY	361	ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTTCCATGAACATATGCGGGGTGCTGCC	420
DB	1250	ACTTAAAGTCTCTTCCCGAAGAAAGCTTCCATGAACATATGCGGGGTGCTGCC	1309
QY	421	ACCCCTGAAGGACAGAGCATGCCACCACTCCGTGAGGTTCAGAGGCCCAAGAGG	480
DB	1310	ACCTCTGAGGACAGAGCATGCCACCACTCCGTGAGGTTCAGAGGCCCAAGAGG	1369
QY	481	TCCTGGGGGAACCATGTGGGTGCGATGGGAGCCGTGAGAGATGAGAGAGGCTTCCCTCC	540
DB	1370	CCCTGAGAGAACCATGTGGGTGCGATGGGAGCCGTGAGAGATGAGAGAGGCTTCCCTCC	1429
QY	541	AAGAGAACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC	600
DB	1430	AAGAGAGCCCGGAGGCTCCAGAGGAAACCCCTCTGAGAGAGAAATGCCAGACCGAGC	1489
QY	601	TGAGAGCTGGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGGAAGAACGA	660
DB	1490	TGAGAGCTGGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGGAAGAACGA	1549
QY	661	GTGACAACAGGTGAAGGCCCAAGACCTGGAAGCTTCTCCCGGCAACCTTTCCGCGCCC	720
DB	1550	ATGACAACAGGTGAAGGCCCTGAAGGCCCGGCTTCTCCCGGCAACCTTTCCACCTCC	1609

```
Qy 721 GGGTGGTGAATCGTGGCAGAGGTGGCCCTGGTGGCATGCCGGGAGGAAGAGTGGCTCAT 780
Db 1610 GGGTGGTGAATCGTGGCAGAGGTGGCCCTGGTGGCATGCCGGGAGGAAGAGTGGCTCAT 1669
Qy 781 GGAATCGTGGTGGTCCCGGTGGAATGTTCAAGAGTGGCCGTGGTGGAGACAGAGTGGCTT 840
Db 1670 GGAACCGTGGTGGTCTCTGGAGGAATGTTCAAGAGTGGCAGAGTGGAGACAGAGAGGCTT 1729
Qy 841 CCGTGGTGGCCCGGGGCAATGGAACGAGGTGGCTTTGGTGGAGGAAGACGAGGTGGCCCTGG 900
Db 1730 CCGAGGTGGCC-GGGAATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAGGTGGTCTGG 1788
Qy 901 GGGGCCCCCTTGAATGGAACAGATGGGAGGAAGAGAGAGAGAGTGGAGGACC 960
Db 1789 GGGGCTTCTTGAATGGAACAGATGGGAGGAAGAGAGAGAGAGAGTGGAGGACC 1848
Qy 961 TGGAAAAATGATTAAGCGAGCAACGTCAGAGAGCGAGAGATGGCCCTTAC 1012
Db 1849 TGGAAAAATGATTAAGCGAGCAACGTCAGAGAGCGAGAGACCGGCCCTTAC 1900
```

Search completed: February 20, 2005, 16:17:55  
Job time : 2994.19 secs

**This Page Blank (uspto)**





CC diseases associated with receptor dysfunction such as prostatic cancer,  
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
 CC testicular feminisation. This method provides reliable, sensitive,  
 CC simple, inexpensive and rapid assessment of the hormonal effects of these  
 CC test compounds. This polypeptide sequence is the human Ewing sarcoma  
 CC protein of the invention.

XX Sequence 656 AA;

Query Match 100.0%; Score 3633; DB 8; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-216;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDVSSTYSQAQAAGQSYAATPQTGAYAGTAAAGQSSYGTGQPTDVSSTYQAQTAT 60  
 DB 1 MASTDVSSTYSQAQAAGQSYAATPQTGAYAGTAAAGQSSYGTGQPTDVSSTYQAQTAT 60  
 QY 61 YGQTAVATSYGQPTGTTTPTAPQAYSQPVQYGTGAYDTTATVTTTQAASVAAQAYGT 120  
 DB 61 YGQTAVATSYGQPTGTTTPTAPQAYSQPVQYGTGAYDTTATVTTTQAASVAAQAYGT 120  
 QY 121 QPAYPAYGQPPAATATPRTDGNKPTTSQPSSTGTGYNPSLQYGSNTSYQVPGSY 180  
 DB 121 QPAYPAYGQPPAATATPRTDGNKPTTSQPSSTGTGYNPSLQYGSNTSYQVPGSY 180  
 QY 181 MQPVTAPSPYPTSYSTPTSTYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 DB 181 MQPVTAPSPYPTSYSTPTSTYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 QY 241 PPTGYSQAAPSYQSSSYGQSSYQONTYGGPSSYGGQSSYGGQSSYGGQSSYGGQPTSY 300  
 DB 241 PPTGYSQAAPSYQSSSYGQSSYQONTYGGPSSYGGQSSYGGQSSYGGQSSYGGQPTSY 300  
 QY 301 GRGGRGFDRCGMSRGGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
 DB 301 GRGGRGFDRCGMSRGGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
 QY 361 SATYVQGLNLSVTLDDLADFPKCGGVVKNMKRTGQPMIHIYLDKRTGKPGDATTVS 420  
 DB 361 SATYVQGLNLSVTLDDLADFPKCGGVVKNMKRTGQPMIHIYLDKRTGKPGDATTVS 420  
 QY 421 PTKAAVWFDFKDFQSKLKVSLARKKPPMNSMRGGLPRREGKMPPLRGGRGGRGGR 480  
 DB 421 PTKAAVWFDFKDFQSKLKVSLARKKPPMNSMRGGLPRREGKMPPLRGGRGGRGGR 480  
 QY 481 GGPWGRMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 DB 481 GGPWGRMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 QY 541 NCKAPRBEGLFPPPPPGDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 DB 541 NCKAPRBEGLFPPPPPGDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 QY 601 GGRGMDRGFGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
 DB 601 GGRGMDRGFGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

RESULT 2  
 ABM82330  
 ID ABM82330 standard; protein; 656 AA.

XX ABM82330;  
 DT 18-NOV-2004 (first entry)  
 XX Tumour-associated antigenic target (TAT) polypeptide PRO58232, SEQ:5990.  
 XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KM central nervous system cancer; bladder cancer; pancreatic cancer;  
 KM cervical cancer; melanoma; leukaemia; hybridisation probe;

KM chromosome identification; chromosome mapping; gene mapping;  
 KM gene therapy; cytostatic.

XX Homo sapiens.

PN WO2004030615-A2.

PD 15-APR-2004.

PF 29-SEP-2003; 2003MO-US028547.

PR 02-OCT-2002; 2002US-0414971P.

PA (GETH) GENENTECH INC.

PI Wu TD, Zhang Z, Zhou Y;

DR WPI; 2004-347921/32.

XX N-PSDB; ACN40903.

PT New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.

PS Claim 12; SEQ ID NO 5990; 7273pp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides, expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT polypeptide of the invention

XX Sequence 656 AA;

Query Match 100.0%; Score 3633; DB 8; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-216;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDVSSTYSQAQAAGQSYAATPQTGAYAGTAAAGQSSYGTGQPTDVSSTYQAQTAT 60  
 DB 1 MASTDVSSTYSQAQAAGQSYAATPQTGAYAGTAAAGQSSYGTGQPTDVSSTYQAQTAT 60  
 QY 61 YGQTAVATSYGQPTGTTTPTAPQAYSQPVQYGTGAYDTTATVTTTQAASVAAQAYGT 120  
 DB 61 YGQTAVATSYGQPTGTTTPTAPQAYSQPVQYGTGAYDTTATVTTTQAASVAAQAYGT 120  
 QY 121 QPAYPAYGQPPAATATPRTDGNKPTTSQPSSTGTGYNPSLQYGSNTSYQVPGSY 180  
 DB 121 QPAYPAYGQPPAATATPRTDGNKPTTSQPSSTGTGYNPSLQYGSNTSYQVPGSY 180  
 QY 181 MQPVTAPSPYPTSYSTPTSTYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 DB 181 MQPVTAPSPYPTSYSTPTSTYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 QY 241 PPTGYSQAAPSYQSSSYGQSSYQONTYGGPSSYGGQSSYGGQSSYGGQSSYGGQPTSY 300  
 DB 241 PPTGYSQAAPSYQSSSYGQSSYQONTYGGPSSYGGQSSYGGQSSYGGQSSYGGQPTSY 300

```

QY 301 GRGCGFDRGMSRGRGCGGMSAGRGFGFNKRGCGPMDDEGPDLDLGPVDPDESDN 360
DB 301 GRGCGFDRGMSRGRGCGGMSAGRGFGFNKRGCGPMDDEGPDLDLGPVDPDESDN 360
QY 361 SAIVYGLNDSVTLDDLADFFKCCGVVKNMKTGTGPMIHIYLDKETGKPRGDAVSYEDP 420
DB 361 SAIVYGLNDSVTLDDLADFFKCCGVVKNMKTGTGPMIHIYLDKETGKPRGDAVSYEDP 420
QY 421 PTKAAVWFVDGDFGSKLKVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGP 480
DB 421 PTKAAVWFVDGDFGSKLKVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGP 480
QY 481 GGPMMGMRGGRGDRGGRGPPRGPRGSRGNPSGGNVQHRAGDMOCPPNGCNGNPFAMRTEC 540
DB 481 GGPMMGMRGGRGDRGGRGPPRGPRGSRGNPSGGNVQHRAGDMOCPPNGCNGNPFAMRTEC 540
QY 541 NCKAKRPGGFLPPPPPGGDRGGRGPGGMRGGRGGLMDRGPGGWFGRGGRGDRGGR 600
DB 541 NCKAKRPGGFLPPPPPGGDRGGRGPGGMRGGRGGLMDRGPGGWFGRGGRGDRGGR 600
QY 601 GGRGMDRGGRGGRGGRGPGGPPGLMEOMGRRGGRGPGKMDKGEHRQERRDRPY 656
DB 601 GGRGMDRGGRGGRGGRGPGGPPGLMEOMGRRGGRGPGKMDKGEHRQERRDRPY 656

RESULT 3
AAR44555
ID AAR44555 standard; protein; 656 AA.
XX
AC AAR44555;
XX
DT 25-MAR-2003 (revised)
DT 26-MAY-1994 (first entry)
XX
DE Human EWS protein deduced from foetal brain cDNA clone B1AC5.
XX
KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;
KW malignant melanoma; human chromosome 11;
KW primitive peripheral neuroectodermal tumour; human chromosome 22;
KW hum-fl1-1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 569
FT /note="corresponds to GGT codon"
XX
PN WO9323549-A2.
XX
PD 25-NOV-1993.
XX
PF 19-MAY-1993; 93WO-FR000494.
XX
PR 20-MAY-1992; 92FR-00006123.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
PI Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougastel B;
PI Thomas G, Zucman J;
DR N-PSDB; AAO50643.
DR WPI; 1993-386580/48.
XX
PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
PT involved in chromosomal translocation, also derived mRNA, probes, fusion
PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
XX
PS Disclosure; Fig 6; 123p; French.
XX
CC The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA
CC library (Stratagene cat. # 936206). The clone B1AC5 was identified and
CC sequenced. It represents the entire coding region and 3'-UTR of the Ews

```

```

CC gene. (updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 656 AA;
Query Match 99.8%; Score 3624; DB 2; Length 656;
Best Local Similarity 99.8%; Pred. No. 4,9e-216;
Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASTDYISYQAQAQGSAYTAQPTQGYAQTTOAYGQGSYGTGYGPTDVSYTOAQTAT 60
DB 1 MASTDYISYQAQAQGSAYTAQPTQGYAQTTOAYGQGSYGTGYGPTDVSYTOAQTAT 60
QY 61 YGCTAATSYGQPTPTTPTAQAISOPVQGGTGAAYTTTATVTTTQASVAAQSAAYGT 120
DB 61 YGCTAATSYGQPTPTTPTAQAISOPVQGGTGAAYTTTATVTTTQASVAAQSAAYGT 120
QY 121 QPAYVAGQOPATAATPRTODGNKPTETSPQSSGTGYNQPSLGYQSNYSYQVPGSY 160
DB 121 QPAYVAGQOPATAATPRTODGNKPTETSPQSSGTGYNQPSLGYQSNYSYQVPGSY 160
QY 181 MQPVTAPSYPTSYSTQPTSYDQSSYQONTYQGPSSYQGSYQGSYQGSYQGPPTSY 240
DB 181 MQPVTAPSYPTSYSTQPTSYDQSSYQONTYQGPSSYQGSYQGSYQGSYQGPPTSY 240
QY 241 PPGTGSYSQAPPSYSGQSSSYGQSSFRQDHPSSMGTGQESGSGFGSGENRSMGPDNR 300
DB 241 PPGTGSYSQAPPSYSGQSSSYGQSSFRQDHPSSMGTGQESGSGFGSGENRSMGPDNR 300
QY 301 GRGCGFDRGMSRGRGCGGMSAGRGFGFNKRGCGPMDDEGPDLDLGPVDPDESDN 360
DB 301 GRGCGFDRGMSRGRGCGGMSAGRGFGFNKRGCGPMDDEGPDLDLGPVDPDESDN 360
QY 361 SAIVYGLNDSVTLDDLADFFKCCGVVKNMKTGTGPMIHIYLDKETGKPRGDAVSYEDP 420
DB 361 SAIVYGLNDSVTLDDLADFFKCCGVVKNMKTGTGPMIHIYLDKETGKPRGDAVSYEDP 420
QY 421 PTKAAVWFVDGDFGSKLKVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGP 480
DB 421 PTKAAVWFVDGDFGSKLKVSLARKKPPMNSMRGGLPREGRGMPPLRGPGGPGGP 480
QY 481 GGPMMGMRGGRGDRGGRGPPRGPRGSRGNPSGGNVQHRAGDMOCPPNGCNGNPFAMRTEC 540
DB 481 GGPMMGMRGGRGDRGGRGPPRGPRGSRGNPSGGNVQHRAGDMOCPPNGCNGNPFAMRTEC 540
QY 541 NCKAKRPGGFLPPPPPGGDRGGRGPGGMRGGRGGLMDRGPGGWFGRGGRGDRGGR 600
DB 541 NCKAKRPGGFLPPPPPGGDRGGRGPGGMRGGRGGLMDRGPGGWFGRGGRGDRGGR 600
QY 601 GGRGMDRGGRGGRGGRGPGGPPGLMEOMGRRGGRGPGKMDKGEHRQERRDRPY 656
DB 601 GGRGMDRGGRGGRGGRGPGGPPGLMEOMGRRGGRGPGKMDKGEHRQERRDRPY 656

RESULT 4
AAM33813
ID AAM33813 standard; protein; 656 AA.
XX
AC AAM33813;
XX
DT 06-JUL-1998 (first entry)
XX
DE Human EWS protein.
XX
KW Tat stimulatory factor; Tat-SF1; transcriptional activity factor; HIV-1;
KW infection; gene therapy; EWS.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 30..44
FT /note="imperfect repeat peptide showing homology to Tat-SF1"
FT Region 209..236

```

PT /note= "Imperfect repeat peptide showing homology to Tat-  
Srl"  
XX MO9800695-A2.  
XX 08-JAN-1998.  
XX 03-JUL-1997; 97MO-US011713.  
XX 03-JUL-1996; 96US-0021218P.  
XX 13-DEC-1996; 96US-0033152P.  
XX (MASI ) MASSACHUSETTS INSTR TECHNOLOGY.  
XX Sharp PA, Zhou Q;  
XX WPI; 1998-087086/08.  
XX  
XX Nucleic acid encoding Tat stimulating factor protein and related  
XX transformed cells - proteins and binding agents, used to treat human  
XX immunodeficiency virus infection.  
XX  
XX Example 6; Page 47-51; 68pp; English.  
XX  
XX EWS is a member of a novel family of putative transcription factors that  
XX have RNA recognition motifs and which are frequently associated with many  
XX types of sarcomas. It shows homology to a novel human transcriptional  
XX activity factor, 'Tat-stimulatory factor' (Tat-SF1, see AAW3811), that is  
XX involved in the regulation of transcriptional elongation of HIV-1 by Tat,  
XX is essential for Tat trans-activation and is a substrate of an associated  
XX cellular kinase. Tat-SF can be used to screen for binding agents useful  
XX in the treatment of HIV infection  
XX  
XX Sequence 656 AA;

Query Match 99.4%; Score 3611; DB 2; Length 656;  
Best Local Similarity 99.7%; Pred. No. 3.1e-215;  
Matches 654; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MASTDSTSTYQAAAGGYSAYTAQPTGTAQTTQAYGQSSYGYGQPTDVSYTAQTTAT 60  
DB 1 MASTDSTSTYQAAAGGYSAYTAQPTGTAQTTQAYGQSSYGYGQPTDVSYTAQTTAT 60  
QY 61 YGQTAATATSGGPTGTTTPTAQAQVAGVQGTGATVDTTATTTTQASVAAQSYGT 120  
DB 61 YGQTAATATSGGPTGTTTPTAQAQVAGVQGTGATVDTTATTTTQASVAAQSYGT 120  
QY 121 QPAYPAAGQPPAATAPTRPDGKPKTETSQPOSSTGYNPSLGYGQSNYSYQVPSGYP 180  
DB 121 QPAYPAAGQPPAATAPTRPDGKPKTETSQPOSSTGYNPSLGYGQSNYSYQVPSGYP 180  
QY 181 MQVYVAPSPPTPSYSTPTSTYDQSSYQONTTGYGQSSYGYGQSSYGYGQPTSTY 240  
DB 181 MQVYVAPSPPTPSYSTPTSTYDQSSYQONTTGYGQSSYGYGQSSYGYGQPTSTY 240  
QY 241 PPTGYSQAPSOYSSQSSSYGQSSYFRDHPSPMGVYGGESGFGSPGNRSMGSGDNR 300  
DB 241 PPTGYSQAPSOYSSQSSSYGQSSYFRDHPSPMGVYGGESGFGSPGNRSMGSGDNR 300  
QY 301 GRRGGRGDRGMRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 360  
DB 301 GRRGGRGDRGMRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 360  
QY 361 SAIVYGGLNSVTLDDLADFPKCGVVKMKRTGQPMIHTIYLDKFGSKPGDAIVSYEDP 420  
DB 361 SAIVYGGLNSVTLDDLADFPKCGVVKMKRTGQPMIHTIYLDKFGSKPGDAIVSYEDP 420  
QY 421 PTKAAVEMFDGKDFGSKGLKVSILARKKPPMNSWRGGLPPRBRGMPPLRGSGGPGGP 480  
DB 421 PTKAAVEMFDGKDFGSKGLKVSILARKKPPMNSWRGGLPPRBRGMPPLRGSGGPGGP 480  
QY 481 GGPMDRGRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 540  
DB 481 GGPMDRGRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 540

DB 481 GGPMDRGRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 540  
QY 541 NOCKAPKPEGFLPPPPPPPGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 600  
DB 541 NOCKAPKPEGFLPPPPPPPGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 600  
QY 601 GGRGMDRGGFGGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 656  
DB 601 GGRGMDRGGFGGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRRGRR 656

# RESULT 5

ABG06460  
ID ABG06460 standard; protein; 665 AA.  
XX  
XX ABG06460;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #6451.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001MO-US008631.  
PF  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-APR-2000; 2000US-00649167.  
XX  
XX (HYSB-) HYSBQ INC.  
PA  
XX  
XX Dmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS70647.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostic, forensic, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
PT  
XX  
XX Claim 20; SEQ ID NO 36819; 103pp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostic, forensic, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
XX amino acid sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 665 AA;

Query Match 99.1%; Score 3600.5; DB 4; Length 665;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-214;  
 Matches 652; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAAGQSYAATAQPTGVAQTT-QAAYGQSSYGTGPTDVSYTAQTTA 59  
 DB 9 MASTDSTYSQAQAAGQSYAATAQPTGVAQTTQAAYGQSSYGTGPTDVSYTAQTTA 68  
 QY 60 TYGQTAATSYGQPTGPTTPTAQAAYGQVGYGTAATTTATTTTQAATAAGAYG 119  
 DB 69 TYGQTAATSYGQPTGPTTPTAQAAYGQVGYGTAATTTATTTTQAATAAGAYG 128  
 QY 120 TQAPAYGQAPATATPTPDGKPTETSDPOSSTGYNPSLGYGQSSYSPQVGSY 179  
 DB 129 TQAPAYGQAPATCTPTEDKPKETISQPOSTGTGYNPSLGYGQSSYSPQVGSY 188  
 QY 180 PMQPTAPSPPTSYSTPTSTYDOSYSQONTYGQPSYGGQSSYGGQSSYGGQPTTS 239  
 DB 189 PMQPTAPSPPTSYSTPTSTYDOSYSQONTYGQPSYGGQSSYGGQSSYGGQPTTS 248  
 QY 240 YPQTGSYQAPEQYSQSSSYGQSSYFRQDHPSSMGVYQESGGFSGPENRSMGPDN 299  
 DB 249 YPQTGSYQAPEQYSQSSSYGQSSYFRQDHPSSMGVYQESGGFSGPENRSMGPDN 308  
 QY 300 RGRGGRGPDGRGSGRGGGAGRGGPNKRGKGMDSGPDLDLGPVDPDESD 359  
 DB 309 RGRGGRGPDGRGSGRGGGAGRGGPNKRGKGMDSGPDLDLGPVDPDESD 368  
 QY 360 NSAIYVQGLNDSTLADLADFPKQCGVKNKKTGQPMIHIYLDKETGPKGDAITYED 419  
 DB 369 NSAIYVQGLNDSTLADLADFPKQCGVKNKKTGQPMIHIYLDKETGPKGDAITYED 428  
 QY 420 PPTAKAAVWFQDGFQSGSLKATSLARKKPPNMSMRGGLPRRGRGMPPLRGSGPGG 479  
 DB 429 PPTAKAAVWFQDGFQSGSLKATSLARKKPPNMSMRGGLPRRGRGMPPLRGSGPGG 488  
 QY 480 PGSPMGRMGGRGDRGFPFRGPRGSGKGNPSGGGNVQHRAGDWCPCPPGCGNPFARTE 539  
 DB 489 PGSPMGRMGGRGDRGFPFRGPRGSGKGNPSGGGNVQHRAGDWCPCPPGCGNPFARTE 548  
 QY 540 CNOCKAPKEGFLPPPPPPGCDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGGRGF 599  
 DB 549 CNOCKAPKEGFLPPPPPPGCDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGGRGF 608  
 QY 600 RGRGGRGDRGGRGGRGPGGPPGLMEQNGRRGGRGGRGKMDKGEHRQERDRPY 656  
 DB 609 RGRGGRGDRGGRGGRGPGGPPGLMEQNGRRGGRGGRGKMDKGEHRQERDRPY 665

RESULT 6  
 ABB57126  
 ID ABB57126 standard; protein; 655 AA.  
 AC ABB57126;  
 XX 07-MAR-2002 (first entry)  
 DE Mouse ischaemic condition related protein sequence SEQ ID NO:289.  
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasoospastic ischaemia; ischaemic condition; ischaemic disease.  
 XX Mus musculus.  
 OS WO200188188-A2.  
 PN 22-NOV-2001.  
 PD 18-MAY-2001; 2001MO-JF004192.  
 PF 18-MAY-2001; 2000JP-00145977.  
 PR 18-MAY-2000; 2000JP-00145977.  
 XX (UYN1-) UNIT NIIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Aaei S, Takahashi Y, Nagata T, Ishii Y;  
 PI WPI; 2002-034733/04.  
 XX N-PSDB; AB199383.  
 DR Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 PS Claim 2; Page 799-802; 2690pp; English.  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (i) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (i). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasoospastic ischaemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. AB199913 and AB199914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention

Sequence 655 AA:  
 QY 1 MASTDSTYSQAQAAGQSYAATAQPTGVAQTTQAAYGQSSYGTGPTDVSYTAQTTA 60  
 DB 1 MASTDSTYSQAQAAGQSYAATAQPTGVAQTTQAAYGQSSYGTGPTDVSYTAQTTA 60  
 QY 61 YGQTAATSYGQPTGPTTPTAQAAYGQVGYGTAATTTATTTTQAATAAGAYG 120  
 DB 61 YGQTAATSYGQPTGPTTPTAQAAYGQVGYGTAATTTATTTTQAATAAGAYG 120  
 QY 121 QPAPYVYGOQPAATATPTPDGKPTETSDPOSSTGYNPSLGYGQSSYSPQVGSY 180  
 DB 121 QPAPYVYGOQPAATATPTPDGKPTETSDPOSSTGYNPSLGYGQSSYSPQVGSY 180  
 QY 181 MOPVTAPEPSYPTSYSTPTSTYDOSYSQONTYGQPSYGGQSSYGGQSSYGGQPTTS 240  
 DB 181 MOPVTAPEPSYPTSYSTPTSTYDOSYSQONTYGQPSYGGQSSYGGQSSYGGQPTTS 240  
 QY 241 PPTGSYQAPEQYSQSSSYGQSSYFRQDHPSSMGVYQESGGFSGPENRSMGPDN 300  
 DB 241 PPTGSYQAPEQYSQSSSYGQSSYFRQDHPSSMGVYQESGGFSGPENRSMGPDN 300  
 QY 301 GGRGGRGPDGRGSGRGGGAGRGGPNKRGKGMDSGPDLDLGPVDPDESD 359  
 DB 301 GGRGGRGPDGRGSGRGGGAGRGGPNKRGKGMDSGPDLDLGPVDPDESD 359  
 QY 361 NSAIYVQGLNDSTLADLADFPKQCGVKNKKTGQPMIHIYLDKETGPKGDAITYED 419  
 DB 361 NSAIYVQGLNDSTLADLADFPKQCGVKNKKTGQPMIHIYLDKETGPKGDAITYED 419  
 QY 421 PPTAKAAVWFQDGFQSGSLKATSLARKKPPNMSMRGGLPRRGRGMPPLRGSGPGG 480  
 DB 421 PPTAKAAVWFQDGFQSGSLKATSLARKKPPNMSMRGGLPRRGRGMPPLRGSGPGG 480  
 QY 481 GGPGRMGGRGDRGFPFRGPRGSGKGNPSGGGNVQHRAGDWCPCPPGCGNPFARTEC 539  
 DB 481 GGPGRMGGRGDRGFPFRGPRGSGKGNPSGGGNVQHRAGDWCPCPPGCGNPFARTEC 539  
 QY 541 NCKKAPKEGFLPPPPPPGCDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGGRGF 600  
 DB 541 NCKKAPKEGFLPPPPPPGCDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGGRGF 600



molecules and the encoded proteins thereof. Specifically, it refers to CDNA clones obtained by an oligo-capping method, where none of these clones are identical to any known human mRNAs. The present invention describes an immunosassay to identify agonists and antagonists, as well as antibodies, antisense molecules and siRNAs that can all be used to bind to and modulate expression of the cDNA molecules. As such, these molecules are useful for diagnostic markers or therapeutic targets for the various diseases or morbid states. In particular, they are useful in gene therapy for treating osteoporosis, neurological disease, Alzheimer's disease, Parkinson's disease, dementia, short memory and various cancers, as well as for maintaining equilibrium of sense or motor function, and for treating emotional reaction, fear response and panic. Accordingly, they exhibit osteoprotective, neuroprotective, nootropic, antiparkinsonian, cytostatic and tranquilizer activities. This polypeptide is a protein encoded by a full length human cDNA sequence of the invention. NOTE: This sequence is not given in the sequence listing of the specification but can be obtained on CD-ROM from the European Patent Office, Vienna Sub-office.

XX Sequence 604 AA;

Query Match 88.4%; Score 3210; DB 8; Length 604;  
Best Local Similarity 89.9%; Pred. No. 1.9e-190;  
Matches 590; Conservative 3; Mismatches 11; Indels 52; Gaps 4;

1 MASTDSTYSGAQAAGCYSATYTAOTTAQYQGSYGYGPTDSTYQAQTAT 60  
1 MASTDSTYSGAQAAGCYSATYTAOTTAQYQGSYGYGPTDSTYQAQTAT 60  
61 YGQTATATSGGPTGTTTAPQAYGQPVQGYGTGATTTATTTTQASVAAQSYGT 120  
61 YGQTATATSGGPTGTTTAPQAYGQPVQGYGTGATTTATTTTQASVAAQSYGT 120  
121 QPAYPAYGQQAATATPRTDGNKPTETSGQSSSTGYNPSLQYGSNYSYQVPSY 180  
100 -----YSGQPAATATPRTDGNKPTETSGQSSSTGYNPSLQYGSNYSYQVPSY 153  
181 MGVTAATPSTYPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 240  
154 MGVTAATPSTYPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 188  
241 PPTGTSYGAQPSYSGQSSSTGYNPSLQYGSNYSYQVPSY 300  
189 PPTGTSYGAQPSYSGQSSSTGYNPSLQYGSNYSYQVPSY 248  
301 GRRGGSFDRGMSRGGRRGMSRGGRRGMSRGGRRGMSRGGRRGMSRGGRRGMS 360  
249 GRRGGSFDRGMSRGGRRGMSRGGRRGMSRGGRRGMSRGGRRGMSRGGRRGMS 308  
361 SATYVGLDSTVLDLADFFKCCGYVKNKRTGQPMIHTYLDKFGPKGATVSYEDP 420  
309 SATYVGLDSTVLDLADFFKCCGYVKNKRTGQPMIHTYLDKFGPKGATVSYEDP 368  
421 PPAKAAVEMPDGDFGSKLKVSLARKKPPMNSMRGGLPRREGMPPLRGSGRGSG 480  
369 PPAKAAVEMPDGDFGSKLKVSLARKKPPMNSMRGGLPRREGMPPLRGSGRGSG 428  
481 GSGMGMGRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 540  
429 GSGMGMGRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 488  
541 NQCKAKPEBFLPPLPPPPGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 600  
489 NQCKAKPEBFLPPLPPPPGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 548  
601 GSGMGMGRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 656  
549 GSGMGMGRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGRGGRRGSGR 604

RESULT 9  
ADP56334  
ID ADP56334 standard; protein; 583 AA.

XX ADP56334;

XX 18-NOV-2004 (first entry)

XX Human PRO protein sequence SEQ ID NO:2310.

XX human; PRO; immune related disease; inflammatory immune response;  
XX immune response stimulation; antiallergic; antianaemic; antiarthritic;  
XX antidiabetic; antidiabetic; antiinflammatory; antipsoriatic;  
XX antineumatic; antithyroid; CNS; dermatological; gastrointestinal;  
XX haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;  
XX nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
XX virucide; gene therapy.

XX Homo sapiens.

XX NC02004039956-A2.

XX 13-MAY-2004.

XX 28-OCT-2003; 2003MO-US034381.

XX 29-OCT-2002; 2002US-0422472P.

XX (GENT ) GENENTECH INC.

XX Agarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;

XX Wood WI, Wu TD;

XX WPI; 2004-376182/35.

XX N-PSDB; ADP56333.

XX New PRO polynucleotides and polypeptides, useful in diagnosing  
XX and treating an immune related disease, e.g. systemic lupus  
XX erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in  
XX stimulating an immune response.

XX Claim 1; SEQ ID NO 2310; 3009pp; English.

XX The present invention describes an isolated PRO nucleic acid (I). Also  
XX described: (1) a vector comprising (1); (2) a host cell comprising the  
XX vector of (1); (3) a process for producing a PRO polypeptide; (4) an  
XX isolated PRO polypeptide; (5) a chimeric molecule comprising the  
XX polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
XX antibody which specifically binds to a polypeptide of (4); (7) a  
XX composition of matter comprising a polypeptide of (4), an agonist or  
XX antagonist of the polypeptide or an antibody that binds to the  
XX polypeptide in combination with a carrier; (8) an article of manufacture  
XX comprising a container, a label on the container and a composition of  
XX matter of (7); (9) a method of treating an immune related disease in a  
XX mammal; (10) a method for determining the presence of a PRO polypeptide  
XX in a sample suspected of having the polypeptide; (11) a method of  
XX diagnosing an immune related disease or an inflammatory immune response  
XX in a mammal; (12) a method of identifying a compound that inhibits or  
XX mimics the activity of or expression of a gene encoding a PRO polypeptide  
XX; and (13) a method of stimulating the immune response in a mammal. The  
XX PRO sequences have antiallergic, antianaemic, antiarthritic,  
XX antineumatic, antidiabetic, antiinflammatory, antipsoriatic,  
XX antirheumatic, antithyroid, CNS, dermatological, gastrointestinal,  
XX haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,  
XX nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and  
XX virucide activities, and can be used in gene therapy. The nucleic acid  
XX (I) and the encoded polypeptides, compositions, kits and methods are  
XX useful in diagnosing and treating an immune related disease and in  
XX stimulating an immune response. The present sequence represents a human  
XX PRO protein from the present invention.

XX Sequence 583 AA;

Query Match 87.5%; Score 3178.5; DB 8; Length 583;  
Best Local Similarity 88.7%; Pred. No. 1.6e-188;  
Matches 582; Conservative 0; Mismatches 1; Indels 73; Gaps 1;



```

QY 1 MASTDSTVSQAAGQSSAATTAQPTGVAOTTAAGQSSGYTGQPTDVSYTOAQTAT 60
DB 1 MASTDSTVSQAAGQSSAATTAQPTGVAOTTAAGQSSGYTGQPTDVSYTOAQTAT 60
QY 61 YGQTAATVSYGQPTGYTPTTAQPAVSQPVQYGTGAVDTTATVTTTQAASVAQSAVGT 120
DB 61 YGQTAATVSYGQPTGYTPTTAQPAVSQPVQYGTGAVDTTATVTTTQAASVAQSAVGT 120
QY 121 QPAVPAATGQPAATAPRRPDGKNTPTTSQPOSSTGTYNPSLGGQSNSTYQVPGSYP 180
DB 121 QPAVPAATGQPAATAPRRPDGKNTPTTSQPOSSTGTYNPSLGGQSNSTYQVPGSYP 180
QY 181 MOPVTAAPSTPTSTVSSTPTSTYDQSSSTQONTYQGSSTYGGQSSYGGQSPPTSY 240
DB 181 MOPVTAAPSTPTSTVSSTPTSTYDQSSSTQONTYQGSSTYGGQSSYGGQSPPTSY 240
QY 241 PPTGTSYQAAPSOYSQQSSSYGQSSSFRQDHPSSMGVYQGESGQSGFSGENRSMGPDNR 300
DB 241 PPTGTSYQAAPSOYSQQSSSYGQSSSFRQDHPSSMGVYQGESGQSGFSGENRSMGPDNR 300
QY 301 GRGRGCGDRGCMRGRGGRGCMGASGERGCFNKPGCGPMDGPDLDLGPVPDDESDN 360
DB 266 -----PMDEGPDLDLGPVPDDESDN 287
QY 361 SAITYGGLANDSVTLDDLADPFKCGVYKMKRRTGQPMIHTYLDKGTGKPKGDAATVSIEDP 420
DB 288 SAITYGGLANDSVTLDDLADPFKCGVYKMKRRTGQPMIHTYLDKGTGKPKGDAATVSIEDP 347
QY 421 PTKAAVWFQKDFQSSKLVSLARKKPPMNSMRGGLPREGRGMPPLRGPGPGPGP 480
DB 348 PTKAAVWFQKDFQSSKLVSLARKKPPMNSMRGGLPREGRGMPPLRGPGPGPGP 407
QY 481 GGPWGRKGGGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
DB 408 GGPWGRKGGGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 467
QY 541 NOCKAPKEBGLPFPFPGDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600
DB 468 NOCKAPKEBGLPFPFPGDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 527
QY 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656
DB 528 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 583

```

RESULT 10  
ABM80197  
ID ABM80197 standard; protein; 361 AA.

```

XX AC ABM80197;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO80700, SEQ:491.
XX XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX KM tumour; diagnosis; cell proliferative disorder; breast cancer;
XX KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX KM central nervous system cancer; bladder cancer; pancreatic cancer;
XX KM cervical cancer; melanoma; leukemia; hybridisation probe;
XX KM chromosome identification; chromosome mapping; gene mapping;
XX KM gene therapy; cytostatic.
XX OS Homo sapiens.
XX PN MO2004030615-A2.
XX PD 15-APR-2004.
XX PF 29-SEP-2003; 2003MO-US028547.
XX PR 02-OCT-2002; 2002US-0414971P.

```

```

XX PA (GETH ) GENENTECH INC.
XX PI Wu TD, Zhang Z, Zhou Y;
XX DR MPI; 2004-347921/32.
XX DR N-PSDB; ACN37537.
XX PT New tumor-associated antigenic target polypeptides and nucleic acids,
XX PT useful in preparing a medicament for treating or detecting a
XX PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX PT prostate cancer or tumor.
XX PS Claim 12; SEQ ID NO 491; 7273bp; English.
XX CC The invention relates to human tumour-associated antigenic target (TAT)
XX CC polypeptides, and their related nucleic acids. The TAT polypeptides are
XX CC overexpressed in cancer tissues compared to normal tissues, and may thus
XX CC serve as effective targets for the diagnosis and treatment of cancer in
XX CC mammals. The invention also relates to nucleic acid and polypeptide
XX CC sequences at least 80% identical to the TAT nucleic acids and
XX CC polypeptides; expression vectors and host cells comprising a TAT nucleic
XX CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX CC TAT polypeptide; and methods and compositions for the treatment or
XX CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX CC antibodies, antagonists, binding molecules and compositions are useful
XX CC for diagnosing or treating a cell proliferative disorder associated with
XX CC increased TAT expression, particularly cancers such as breast cancer,
XX CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX CC cancer, pancreatic cancer, cervical cancer, cancers of the central
XX CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX CC used as hybridisation probes, in chromosome and gene mapping, in
XX CC chromosome identification and in gene therapy. The present sequence
XX CC represents a TAT polypeptide of the invention
XX SQ
XX Sequence 361 AA;

```

Query Match 45.1%; Score 1639; DB 8; Length 361;  
Best Local Similarity 80.2%; Pred. No. 1.9e-93;  
Matches 303; Conservative 14; Mismatches 43; Indels 18; Gaps 2;

```

QY 275 MGVYQGESGAFSGPGENRSMGPDNRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 334
DB 1 MGVYQGESGAFSGPGENRSMGPDNRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 60
QY 335 KPGGPMDEGPDLDLGPVPDDESDNSAITYGGLANDSVTLDDLADPFKCGVYKMKRRTG 394
DB 61 KPGGPMDEGPDLDLGPVPDDESDNSAITYGGLANDSVTLDDLADPFKCGVYKMKRRTG 120
QY 395 QPMIHTYLDKGTGKPKGDAATVSIEDPPTAKAAVWFQKDFQSSKLVSLARKKPPMNSM 454
DB 121 QPMIHTYLDKGTGKPKGDAATVSIEDPPTAKAAVWFQKDFQSSKLVSLARKKPPMNSM 180
QY 455 RGLPFRGGRGMPPLRGPGGPGGPGGPMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 514
DB 181 RGLPFRGGRGMPPLRGPGGPGGPGGPMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 240
QY 515 VOHRAGDMQCPNPGCNQNFAMTEGNCCKAPBEGFLPFPFPGDGRGGRGGRGGRGGR 574
DB 241 VOHRAGDMQCPNPGCNQNFAMTEGNCCKAPBEGFLPFPFPGDGRGGRGGRGGRGGR 300
QY 575 RGLPFRGGRGMPPLRGPGGPGGPGGPMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 634
DB 301 RGLPFRGGRGMPPLRGPGGPGGPGGPMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 342
QY 635 GRGPGKMDKGEHROBR 652
DB 343 ALGGPRTFVTNGRKR 360

```

RESULT 11  
ABG95068

ID ABG95068 standard; protein; 476 AA.  
 XX ABG95068;  
 AC  
 DT 04-DEC-2002 (first entry)  
 DE Human translocation (11; 22) (q24, q12) protein #1.  
 XX  
 XX Chromosome aberration; oncogenic fusion protein; cancer;  
 XX proliferative disease; cellular protein isoform; heat shock protein 90;  
 XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 XX rhabdomyosarcoma; synovial sarcoma; viral infection.  
 OS Homo sapiens.  
 XX  
 XX MO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002MO-US006518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FJ;  
 XX  
 DR WPI, 2002-698710/75.  
 XX N-PSDB; ABS73252.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 DR  
 XX  
 PS Disclosure; Page 204-206; 389pp; English.  
 XX  
 XX The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents a protein encoded by the DNA sequence of a  
 CC chromosome aberration  
 CC  
 XX  
 XX Sequence 476 AA;  
 SQ  
 Query Match 39.3%; Score 1429; DB 5; Length 476;  
 Best Local Similarity 88.8%; Pred. No. 2,6e-80;  
 Matches 270; Conservative 6; Mismatches 16; Indels 12; Gaps 1;  
 QY 1 MASTDVTSTYGAQAAGQSYATTAQPTGVAQTTOAVGQSGSYGYGQPTDVSYTAQQTAT 60  
 DB 1 MASTDVTSTYGAQAAGQSYATTAQPTGVAQTTOAVGQSGSYGYGQPTDVSYTAQQTAT 60  
 QY 61 YGQTAATATSTGCGPPTGTTTPTAPQAVSQPVQVGTGATVDTTATVTTTQASTAQAAGAT 120

DB 61 YGQTAATATSTGCGPPTGTTTPTAPQAVSQPVQVGTGATVDTTATVTTTQASTAQAAGAT 120  
 QY 121 QPAVPAVGQPPAATATPPTPDGKPKPTSTQPOSSGTGYNPSTLGYGSNYSYQVPQPSYP 180  
 DB 121 QPAVPAVGQPPAATATPPTPDGKPKPTSTQPOSSGTGYNPSTLGYGSNYSYQVPQPSYP 180  
 QY 181 MCPVTAPPSYPPSTSYSTQPTSTYDQSSYSQONTTGGPSSYGGQSSYGGQPPSTSY 240  
 DB 181 MCPVTAPPSYPPSTSYSTQPTSTYDQSSYSQONTTGGPSSYGGQSSYGGQPPSTSY 240  
 QY 241 PPQTGSYSAQPSQYSSQSSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 288  
 DB 241 PPQTGSYSAQPSQYSSQSSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 300  
 QY 289 GGNR 292  
 DB 301 GSGQ 304  
 RESULT 12  
 AAM34972  
 ID AAM34972 standard; protein; 362 AA.  
 XX  
 XX AAM34972;  
 AC  
 XX  
 DT 09-MAR-1998 (first entry)  
 XX  
 DE Chimeric Ewing's sarcoma-WT1 protein splice variant 2.  
 XX  
 XX Ewing's sarcoma; EMS; EMS-WT1 protein; peripheral neuroectodermal tumour;  
 XX PNET; breakpoint locus; Wilms' tumour;  
 XX desmoplastic small round cell tumour; DSRC tumour.  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..264  
 FT /note="EMS protein"  
 FT 265..362  
 FT /note="WT1 protein"  
 XX  
 PN US5670317-A.  
 XX  
 PD 23-SEP-1997.  
 XX  
 PF 08-MAY-1995; 95US-00437027.  
 XX  
 PR 08-MAY-1995; 95US-00437027.  
 XX  
 PA (SLOK) SLOAN KETTERING INST CANCER RES.  
 XX  
 PI Ladanyi M, Gerald W;  
 XX  
 DR WPI, 1997-479448/44.  
 XX N-PSDB; AAT97870.  
 XX  
 PT Diagnosis of desmoplastic small round cell tumours - by detecting nucleic  
 PT acid encoding chimeric EMS-WT1 protein.  
 XX  
 PS Disclosure; Col 39-42; 34pp; English.  
 XX  
 XX The present sequence represents a chimeric human Ewing's sarcoma (EMS) -  
 CC WT1 protein. EMS is also known as peripheral neuroectodermal tumour  
 CC (PNET). The protein is composed of EMS and WT1 proteins (see features  
 CC table). The WT1 gene encodes 4 alternatively spliced RNA transcripts that  
 CC produce functionally different products. One alternative splice site is  
 CC at the 3' end of exon 9 and leads to products that differ by 3 amino  
 CC acids placed between the third and fourth zinc fingers of the DNA binding  
 CC domain. The present protein does not contain these three amino acids  
 CC (Lys, Thr, Ser), while AAM34971 does. Detection of EMS-WT1 can be  
 CC accomplished by reverse transcriptase PCR, where WT1 is screened (see  
 CC AAT97852-68) as a breakpoint locus because of its involvement in Wilms'

CC tumour, which shares some histopathologic features of desmoplastic small  
 CC round cell (DSRC) tumours. The EMS-WT1 chimeric transcript has been  
 CC detected in 11 out of 12 DSRC tumours studied and in none of 49 other  
 CC tumours. DSRC tumours are associated with translocation of the EMS gene.  
 CC Oligonucleotides based on the AAT97869 are used in a method for the  
 CC diagnosis of DSRC tumours in patients. The method comprises detecting a  
 CC nucleic acid molecule encoding a chimeric EMS-WT1 protein in a sample  
 CC from the subject, where positive detection indicates the presence of a  
 CC DSRC tumour

XX Sequence 362 AA;

Query Match 39.1%; Score 1420; DB 2; Length 362;

Best Local Similarity 100.0%; Pred. No. 6,9e-80; Mismatches 0; Gaps 0;

Matches 265; Conservative 0; Indels 0; Gaps 0;

QY 1 MASTDYSYTSQAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
 DB 1 MASTDYSYTSQAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60

QY 61 YGOTAVATSYGQPTGTTTTPAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 DB 61 YGOTAVATSYGQPTGTTTTPAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120

QY 121 QPAYPAYGQGPAPATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSYP 180  
 DB 121 QPAYPAYGQGPAPATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSYP 180

QY 181 MQPVTAAPSYPPTSYSTQPTSTYDQSSYQONTYGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
 DB 181 MQPVTAAPSYPPTSYSTQPTSTYDQSSYQONTYGPSSYGGQSSYGGQSSYGGQPTPTSY 240

QY 241 PPOTGSYSQAPSOYSQSSSSYGGQSS 265  
 DB 241 PPOTGSYSQAPSOYSQSSSSYGGQSS 265

#### RESULT 13

AAW34971 standard; protein; 365 AA.

AC AAW34971;

DT 09-MAR-1998 (first entry)

XX Chimeric Ewing's sarcoma-WT1 protein splice variant 1.

XX Ewing's sarcoma; EMS; EMS-WT1 protein; peripheral neuroectodermal tumour;

XX PMBT; breakpoint locus; Wilms' tumour;

XX desmoplastic small round cell tumour; DSRC tumour.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Protein 1..264

FT Protein /note= "EMS protein"

FT Protein 265..365

FT Protein /note= "WT1 protein"

XX US5670317-A.

XX 23-SEP-1997.

XX 08-MAY-1995; 95US-00437027.

XX 08-MAY-1995; 95US-00437027.

XX (SLOK ) SLOAN KETTERING INST CANCER RES.

XX Ladanyi M, Gerald W;

XX WPI; 1997-47948/44.

XX N-PSDB; AAT97869.

XX Diagnosis of desmoplastic small round cell tumours - by detecting nucleic  
 PT acid encoding chimeric EMS-WT1 protein.

XX Disclosure; Col 37-40; 34pp; English.

XX The present sequence represents a chimeric human Ewing's sarcoma (EMS) -  
 CC WT1 protein. EMS is also known as peripheral neuroectodermal tumor  
 CC (PNET). The protein is composed of EMS and WT1 proteins (see features  
 CC table). The WT1 gene encodes 4 alternatively spliced RNA transcripts that  
 CC produce functionally different products. One alternative splice site is  
 CC at the 3' end of exon 9 and leads to products that differ by 3 amino  
 CC acids placed between the third and fourth zinc fingers of the DNA binding  
 CC domain. The present protein contains these three amino acids (Lys, Thr,  
 CC Ser), while AAW34972 does not. Detection of EMS-WT1 can be accomplished  
 CC by reverse transcriptase PCR, where WT1 is screened (see AAT97852-68) as  
 CC a breakpoint locus because of its involvement in Wilms' tumour, which  
 CC shares some histopathologic features of desmoplastic small round cell  
 CC (DSRC) tumours. The EMS-WT1 chimeric transcript has been detected in 11  
 CC out of 12 DSRC tumours studied and in none of 49 other tumours. DSRC  
 CC tumours are associated with translocation of the EMS gene.  
 CC Oligonucleotides based on the AAT97869 are used in a method for the  
 CC diagnosis of DSRC tumours in patients. The method comprises detecting a  
 CC nucleic acid molecule encoding a chimeric EMS-WT1 protein in a sample  
 CC from the subject, where positive detection indicates the presence of a  
 CC DSRC tumour

XX Sequence 365 AA;

Query Match 39.1%; Score 1420; DB 2; Length 365;

Best Local Similarity 100.0%; Pred. No. 6,9e-80; Mismatches 0; Indels 0; Gaps 0;

Matches 265; Conservative 0; Indels 0; Gaps 0;

QY 1 MASTDYSYTSQAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
 DB 1 MASTDYSYTSQAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60

QY 61 YGOTAVATSYGQPTGTTTTPAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 DB 61 YGOTAVATSYGQPTGTTTTPAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120

QY 121 QPAYPAYGQGPAPATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSYP 180  
 DB 121 QPAYPAYGQGPAPATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSYP 180

QY 181 MQPVTAAPSYPPTSYSTQPTSTYDQSSYQONTYGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
 DB 181 MQPVTAAPSYPPTSYSTQPTSTYDQSSYQONTYGPSSYGGQSSYGGQSSYGGQPTPTSY 240

QY 241 PPOTGSYSQAPSOYSQSSSSYGGQSS 265  
 DB 241 PPOTGSYSQAPSOYSQSSSSYGGQSS 265

#### RESULT 14

AD126113 standard; protein; 518 AA.

AC AD126113;

DT 22-APR-2004 (first entry)

XX Human protein that promotes STAT6 activation #39.

XX human; signal transducer and activator of transcription 6; STAT6;

XX immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;

XX diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;

XX rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;

XX sepsis; asthma; allergic rhinitis; ischaemic heart disease;

XX subarachnoid haemorrhage; viral hepatitis; AIDS.

XX Homo sapiens.

PN WO2003104277-A2.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 05-JUN-2003; 2003WO-JP007123.  
 XX  
 PR 05-JUN-2002; 2002JP-00164257.  
 XX  
 PR 06-JUN-2002; 2002JUS-0385912P.  
 PR 26-DEC-2002; 2002JP-00377326.  
 PR 27-DEC-2002; 2002JUS-0436467P.  
 PR 15-MAY-2003; 2003JP-00137505.  
 PR 16-MAY-2003; 2003JUS-0470836P.  
 XX  
 PA (ASAH) ASAH KASEI KK.  
 XX  
 P1 Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
 DR N-PSDB; AD126112.  
 XX  
 DR WPI; 2004-122214/12.  
 XX  
 DR N-PSDB; AD126112.  
 XX  
 PT New signal transducer and activator of transcription 6 activation  
 PT promoting purified protein, for diagnosing and treating disease  
 PT associated with activation/inhibition of transcription factor e.g.  
 PT diabetes and cancer.  
 XX  
 PS Claim 1, SEQ ID NO 78; 1368bp; English.  
 XX  
 CC The invention relates to a purified protein promoting signal transducer  
 CC and activator of transcription 6 activation (STAT6). The protein is  
 CC useful for the producing an antibody, which involves administering the  
 CC protein or its epitope-bearing fragments to a non-human animal as an  
 CC antigen. The nucleic acid is useful for diagnosing a disease or  
 CC susceptibility to a disease related to expression or activity of the  
 CC protein. A transformant expressing the protein is useful for screening  
 CC compounds which inhibit or promote STAT6 activation. A transformant  
 CC expressing the protein is useful for producing a pharmaceutical  
 CC composition. Compositions, antibodies and antisense molecules are useful  
 CC for the treating a disease associated with STAT6 activation such as  
 CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
 CC hyperlipidaemia, infectious disease and cancers. Compositions are useful  
 CC for treating disease associated with STAT6 activation and/or prevention  
 CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
 CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
 CC allergic rhinitis, ischaemic heart diseases, subarachnoid haemorrhage,  
 CC viral hepatitis and AIDS. The protein has efficient promoting STAT6  
 CC activity. The protein or nucleic acid is effectively useful for screening  
 CC compounds for treating and preventing disease associated with excessive  
 CC activation or inhibition of STAT6. The present sequence represents the  
 CC amino acid sequence of a human protein which promotes STAT6 activation.  
 XX  
 SQ Sequence 518 AA;  
 Query Match 35.5%; Score 1290; DB 8; Length 518;  
 Best Local Similarity 43.9%; Pred. No. 1,1e-71;  
 Matches 301; Conservative 59; Mismatches 128; Indels 198; Gaps 30;  
 QY 1 MASTDYSYSAQAAGQYSAHTAQTGVA-QTTQAYGQGSYGTGPGPTDVSYTAQCTTA 59  
 DB 1 MASNDDY-----QATGSGAYPTQPGQYSGQSPVGGQSYSGSAD-----TS 48  
 QY 60 TYGQTAYATSYGQP-IGYTTTPAPVAPVPGYGTGAYDTTATVTTTQASIAAQA 117  
 DB 49 GYQSSSYSGSYGTQNTGYSQASAP-----QYGTGTYG-----SSQSSQSS 91  
 QY 118 YGTQAPAYAGQGPATATAPRPOGKNTENSPQSGTGGNPSLGGQSNYSYVPVPG 177  
 DB 92 YGQSSSYFGIGQGP-----PSSYSGY-----G 116  
 QY 178 SYMPQVPTAPPSYPTSYSTOPTSYDQSSYQNTYGPSS--YQSSSYG-QQSSSYGQ 234  
 DB 117 S-----SGSSSYGQPGQSGYGGQSGYGGQSGYGGQSGYGGQSGYGGQSGYGGQ 146  
 QY 235 QPPTSYPTQGSYQABSYQSSSYGQSSFRQDHPSPKMGVYGGSGGFGSPGGRKSM 294

Db 147 QSSSYNPQ--GYGQ-QNQNSSSGGGGGG-----GNYGDDQSSMSGGGGG 194  
 QY 295 SGPDNRGRGRGF-----DRGMSRG-----GGGGRGAGSAG--ERG 332  
 Db 195 GNQDQSGGGGGYGGGQDQGRGGRGGGGYNNSSGGEYEPGRGGGGGGGSGDRG 254  
 QY 333 FNRGPMGEPPLDLPVPDPEDSDNSAIYQGLNDSTLDDLDLDFQCCVYKNNK 392  
 Db 255 FNRGPMGEPPLDLPVPDPEDSDNSAIYQGLNDSTLDDLDLDFQCCVYKNNK 309  
 QY 393 TQPMTHIYDKETGPKGATATSYEDPPTAKAAYWFDKDPQGSKLAKKPPMN 452  
 Db 310 TQPMTHIYDKETGPKGATATSYEDPPTAKAAYWFDKDPQGSKLAKKPPMN 369  
 QY 453 SMRGLPPEBRGMPPLRLGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 512  
 Db 370 --RGGNGRGRG-----RGPMRGYGGG--SGGGGGRGPFSGG-----GG 409  
 QY 513 GNYQRAQDMQCNPCGQNFAMRTCCNCKAPKTEGFLPPPPPPGGRGGRGPGMR 572  
 Db 410 GGGQQRAGMKCPNPTCEMNFMRBECNCKAPKPDG--PGGPGGSHMGANTGDDR 465  
 QY 573 GGRGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 631  
 Db 466 RGRGG-YDRGG---YR-GRGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 498  
 QY 632 RRGGRGPGMRD-KGHRQERRDRPY 656  
 Db 499 -----GPGMRDGRGHRDRRERY 518  
 RESULT 15  
 AAM78355  
 ID AAM78355 standard; protein; 526 AA.  
 XX  
 AC AAM78355;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1017.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51488.



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 14:59:05 ; Search time 47 Seconds  
(without alignments)  
1041.910 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDTSTYSQAAAQGGYSA.....GGFGKMDKGEHQRDRDPY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	656	2 US-08-343-443B-2	Sequence 2, Appl
2	3633	100.0	656	3 US-09-214-564A-4	Sequence 4, Appl
3	3633	100.0	656	4 US-09-538-092-1250	Sequence 1250, Ap
4	3178.5	87.5	591	4 US-09-949-016-10914	Sequence 10914, A
5	3178.5	87.5	591	4 US-09-949-016-10915	Sequence 10915, A
6	1420	39.1	362	1 US-08-437-027-21	Sequence 21, Appl
7	1420	39.1	365	1 US-08-437-027-20	Sequence 20, Appl
8	1337	36.8	306	4 US-09-949-016-7825	Sequence 7825, Ap
9	1282	35.3	526	4 US-09-538-092-1080	Sequence 1080, Ap
10	484	13.3	295	2 US-08-343-443B-107	Sequence 107, App
11	470.5	13.0	462	4 US-09-919-039-324	Sequence 324, App
12	453	12.5	86	2 US-08-343-443B-7	Sequence 7, Appl
13	376.5	10.4	1078	3 US-08-963-825-21	Sequence 21, Appl
14	376.5	10.4	1078	3 US-09-500-811-21	Sequence 21, Appl
15	376.5	10.4	1078	3 US-09-570-573-21	Sequence 21, Appl
16	376.5	10.4	1078	3 US-09-548-608-21	Sequence 21, Appl
17	373	10.3	1057	3 US-08-931-820-4	Sequence 4, Appl
18	344	9.5	745	2 US-09-010-928B-28	Sequence 28, Appl
19	344	9.5	870	2 US-09-010-928B-4	Sequence 4, Appl
20	339.5	9.3	907	2 US-09-010-928B-2	Sequence 2, Appl
21	338.5	9.3	1065	1 US-08-642-255-72	Sequence 72, Appl
22	337.5	9.3	1461	4 US-09-585-887-9	Sequence 9, Appl
23	337.5	9.3	1461	4 US-09-289-578-9	Sequence 9, Appl
24	335.5	9.2	1418	3 US-08-963-825-20	Sequence 20, Appl
25	335.5	9.2	1418	3 US-09-010-999-1	Sequence 1, Appl
26	335.5	9.2	1418	3 US-09-500-811-20	Sequence 20, Appl
27	335.5	9.2	1418	3 US-09-570-573-20	Sequence 20, Appl

28	335.5	9.2	1418	3 US-09-548-608-20	Sequence 20, Appl
29	334.5	9.2	633	1 US-08-642-255-73	Sequence 73, Appl
30	334	9.2	1739	4 US-09-795-061-2	Sequence 2, Appl
31	333.5	9.2	714	3 US-08-556-978B-61	Sequence 61, Appl
32	333.5	9.2	714	3 US-09-247-806-10	Sequence 10, Appl
33	333	9.2	1464	4 US-09-331-347C-21	Sequence 21, Appl
34	331	9.1	1057	3 US-08-931-820-1	Sequence 1, Appl
35	331	9.1	1745	4 US-09-795-061-4	Sequence 4, Appl
36	330	9.1	1806	4 US-09-919-497-56	Sequence 56, Appl
37	329	9.1	822	3 US-09-219-849-49	Sequence 49, Appl
38	327.5	9.0	1060	3 US-08-931-820-3	Sequence 3, Appl
39	327	9.0	492	4 US-08-468-996-12	Sequence 12, Appl
40	326.5	8.9	1017	4 US-08-468-996-10	Sequence 10, Appl
41	324.5	8.9	1341	3 US-08-963-825-18	Sequence 18, Appl
42	324.5	8.9	1341	3 US-09-500-811-18	Sequence 18, Appl
43	324.5	8.9	1341	3 US-09-570-573-18	Sequence 18, Appl
44	324.5	8.9	1341	3 US-09-548-608-18	Sequence 18, Appl
45	322	8.9	1179	4 US-09-949-016-7088	Sequence 7088, Ap

#### ALIGNMENTS

RESULT 1  
US-08-343-443B-2  
Sequence 2, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Petet, Martine  
APPLICANT: Ploougaestel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8363  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-343-443B-2

```

Query Match	100.0%	Score 3633	DB 2	Length 656
Best Local Similarity	100.0%	Pred. No. 8.4e-243		
Matches 656; Conservative	0	Mismatches	0	Indels 0; Gaps 0

Qy	MASTDYSTYSQAAAQOSYSAYTAQNPQGVYQQTQAQGGOSYGVGGPDTVSTQAQTTAT	60
Db	1 MASTDYSTYSQAAAQOSYSAYTAQNPQGVYQQTQAQGGOSYGVGGPDTVSTQAQTTAT	60
Qy	YQGTAVATSYGQPEPTYTTPTAPOAYSQPVQSYGTGAYDTTAAVTTTTAASYAASAYGT	12
Db	61 YQGTAVATSYGQPEPTYTTPTAPOAYSQPVQSYGTGAYDTTAAVTTTTAASYAASAYGT	12
Qy	QPAYPAVQGPAPATAPRPPDGNKKPRTETSPQSTSTGTGNQPSLIGYQSNYSIPQVPGSY	18
Db	121 QPAYPAVQGPAPATAPRPPDGNKKPRTETSPQSTSTGTGNQPSLIGYQSNYSIPQVPGSY	18
Qy	181 MQPVTAPPSYPTPSTYSSTQPTSYDQSSYSQGNQYQGPSSYQGOOSYSGOOSYSGOOPPTSY	24
Db	181 MQPVTAPPSYPTPSTYSSTQPTSYDQSSYSQGNQYQGPSSYQGOOSYSGOOSYSGOOPPTSY	24
Qy	241 PPTQTSYSQAPPSQYSQOOSYSYGOQSSFRDHPSSMCVYQGESGFBGPEBNSMSGPDR	30
Db	241 PPTQTSYSQAPPSQYSQOOSYSYGOQSSFRDHPSSMCVYQGESGFBGPEBNSMSGPDR	30
Qy	301 GRGCGGDRGMSYSGRGGRGGMGSAGBRGCGNKKRGSGPMDGSPDLIDLAPVPDPEDSN	36
Db	301 GRGCGGDRGMSYSGRGGRGGMGSAGBRGCGNKKRGSGPMDGSPDLIDLAPVPDPEDSN	36
Qy	361 SAIYVQGLANDSVTLDDLADFPKQCGVYKONKRTGQPMIHYLDKETGKEKGDATVSYEDP	42
Db	361 SAIYVQGLANDSVTLDDLADFPKQCGVYKONKRTGQPMIHYLDKETGKEKGDATVSYEDP	42
Qy	421 PTAKAAABEMDGDPOSSKTKVSLAKKPPMNSMRGGLPPRBGRGMPPLRLCGRPGRGPG	48
Db	421 PTAKAAABEMDGDPOSSKTKVSLAKKPPMNSMRGGLPPRBGRGMPPLRLCGRPGRGPG	48
Qy	481 GGPBGRNGRGGRGGRGPPRGRPGSGSGNPGSGGYNVHRAGDMQCPMPGCGMGNFAMRTBC	54
Db	481 GGPBGRNGRGGRGGRGPPRGRPGSGSGNPGSGGYNVHRAGDMQCPMPGCGMGNFAMRTBC	54
Qy	541 NOCKAPVPBGFLLPPPPPPCGDNRGGGPGGMRGGRGGLMDRGPGGMPFRGSGRGGGR	60
Db	541 NOCKAPVPBGFLLPPPPPPCGDNRGGGPGGMRGGRGGLMDRGPGGMPFRGSGRGGGR	60
Qy	601 GGRGMDGSGGGRGGRGPGPPRLPLMOWGGRGCGGGRGMDKGBHREBDRDY	66
Db	601 GGRGMDGSGGGRGGRGPGPPRLPLMOWGGRGCGGGRGMDKGBHREBDRDY	66

RESULT 2  
US-09-214-564A-4  
; Sequence 4, Application US/09214564A

GENERAL INFORMATION: Philip A.  
APPLICANT: Sharp, Qiang  
APPLICANT: Zhou, Qiang  
TITLE OR INVENTION: Tat-SF: Cofactor For Stimulation Of Transcriptional  
TITLE OR INVENTION: Elongation By HIV-1 Tat  
FILE REFERENCE: M0656/7042  
CURRENT APPLICATION NUMBER: US/09/214,564A  
CURRENT FILING DATE: 1999-08-18  
PRIOR APPLICATION NUMBER: US 60/021,218  
PRIOR FILING DATE: 1996-07-03  
PRIOR APPLICATION NUMBER: US 60/033,152  
PRIOR FILING DATE: 1996-12-13  
PRIOR APPLICATION NUMBER: PCT/US97/11713  
PRIOR FILING DATE: 1997-07-03

```

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

```

Query Match	100.0%	Score 3633	DB 3	Length 656
Best Local Similarity	100.0%	Pred. No. 8.4e-243		
Matches 656	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	1	MASTDI	STYS	QAAAQOQGS	ATATAC	PTGCACTT	AAVAGGOS	YCTY	GCP	TV	SV	QAQ	TTAT	60
Db	1	MASTDI	STYS	QAAAQOQGS	ATATAC	PTGCACTT	AAVAGGOS	YCTY	GCP	TV	SV	QAQ	TTAT	60
QY	61	YQGTAA	VA	TSY	QOP	PTGY	TTPTA	PAQ	VSQ	PVQ	GY	GTAY	DTTAT	120
Db	61	YQGTAA	VA	TSY	QOP	PTGY	TTPTA	PAQ	VSQ	PVQ	GY	GTAY	DTTAT	120
QY	121	QPA	PAY	QOQ	PA	ATP	TR	PO	GN	K	TE	T	S	180
Db	121	QPA	PAY	QOQ	PA	ATP	TR	PO	GN	K	TE	T	S	180
QY	181	MQP	TA	PSY	P	T	S	T	S	T	O	P	T	240
Db	181	MQP	TA	PSY	P	T	S	T	S	T	O	P	T	240
QY	241	PPOT	GS	YQ	AP	S	YQ	QO	SS	YQ	QO	SS	YQ	300
Db	241	PPOT	GS	YQ	AP	S	YQ	QO	SS	YQ	QO	SS	YQ	300
QY	301	GRG	RG	FP	RG	MS	RG	RG	CG	MS	AA	BER	GF	360
Db	301	GRG	RG	FP	RG	MS	RG	RG	CG	MS	AA	BER	GF	360
QY	361	SAI	YQ	G	L	N	D	S	V	T	L	L	D	420
Db	361	SAI	YQ	G	L	N	D	S	V	T	L	L	D	420
QY	421	PTAA	RA	V	M	F	P	G	K	O	P	O	G	480
Db	421	PTAA	RA	V	M	F	P	G	K	O	P	O	G	480
QY	481	GGP	M	G	M	G	R	G	D	G	F	P	R	540
Db	481	GGP	M	G	M	G	R	G	D	G	F	P	R	540
QY	541	NOCA	P	K	E	G	F	L	P	P	F	P	P	600
Db	541	NOCA	P	K	E	G	F	L	P	P	F	P	P	600
QY	601	GGR	G	M	D	R	G	F	G	G	R	G	G	656
Db	601	GGR	G	M	D	R	G	F	G	G	R	G	G	656

RESULT 3  
US-09-538-092-1250  
; Sequence 1250, Application US/09538092

1 APPLICANT: Glot, Iolc  
2 APPLICANT: Manefield, Traci A.  
3 TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
4 FILE REFERENCE: 15966-542  
5 CURRENT APPLICATION NUMBER: US/09/538,092  
6 CURRENT FILING DATE: 2000-03-29  
7 PRIOR APPLICATION NUMBER: 60/127,352  
8 PRIOR FILING DATE: 1999-04-01  
9 PRIOR APPLICATION NUMBER: 60/178,965  
10 PRIOR FILING DATE: 2000-02-01  
11 NUMBER OF SEQ ID NOS: 1387



SOFTWARE: CuraPatSeqFormatter Version 0.9

SEQ ID NO 1250  
LENGTH: 656  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (0)...(0)  
OTHER INFORMATION: Polypeptide Accession Number Q01844  
US-09-538-092-1250

Query Match 100.0%; Score 3633; DB 4; Length 656;  
Best Local Similarity 100.0%; Pred. No. 8.4e-243;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQSSYGTGYGPTDVSYTOAQTAT 60  
DB 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQSSYGTGYGPTDVSYTOAQTAT 60  
QY 61 YGQTAATSYGQPTGTTTPTAPQAVSQPVQGYGTGAVDTTTAVTTTQASVAAQSAVGT 120  
DB 61 YGQTAATSYGQPTGTTTPTAPQAVSQPVQGYGTGAVDTTTAVTTTQASVAAQSAVGT 120  
QY 121 QPAPYAGQOPATAATPAPDGNKPTETSQPSSGTGYNQPSLGYGQSNYSYPOVPSYP 180  
DB 121 QPAPYAGQOPATAATPAPDGNKPTETSQPSSGTGYNQPSLGYGQSNYSYPOVPSYP 180  
QY 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240  
DB 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240  
QY 241 PPTGTSYQAAPSOYSQOSSSYGOQSSYFRQDHPSSMGVYQGESGFGSGGENRSMGSDNR 300  
DB 241 PPTGTSYQAAPSOYSQOSSSYGOQSSYFRQDHPSSMGVYQGESGFGSGGENRSMGSDNR 300  
QY 301 GRRGGRGDRGMSRGRGGRGMSAGRGFPKPGPMDDEGDDLDLGPVPDDESDN 360  
DB 301 GRRGGRGDRGMSRGRGGRGMSAGRGFPKPGPMDDEGDDLDLGPVPDDESDN 360  
QY 361 SAIYVQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDEKTKPKGDAVSYEDP 420  
DB 361 SAIYVQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDEKTKPKGDAVSYEDP 420  
QY 421 PTKAAAVEMFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGP 480  
DB 421 PTKAAAVEMFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGP 480  
QY 481 GGPMDRGRGGRGGRGGRGPRGSRGNPSGGGVQHRADWQCPNPGCGNPNFARTTEC 540  
DB 481 GGPMDRGRGGRGGRGGRGPRGSRGNPSGGGVQHRADWQCPNPGCGNPNFARTTEC 540  
QY 541 NCKAPKPEBGLPPLPPPGDGRGGRGGRGGLMDRGPGGMFRGGRGGRGGRGGR 600  
DB 541 NCKAPKPEBGLPPLPPPGDGRGGRGGRGGLMDRGPGGMFRGGRGGRGGRGGR 600  
QY 601 GGRGMDRGRGGRGGRGGRGPGPLMEQMGRRGRGGRGPGKMDKGHRQRRDRPY 656  
DB 601 GGRGMDRGRGGRGGRGGRGPGPLMEQMGRRGRGGRGPGKMDKGHRQRRDRPY 656

RESULT 4  
US-09-949-016-10914

Sequence 10914, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10914  
LENGTH: 591  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-10914

Query Match 87.5%; Score 3178.5; DB 4; Length 591;  
Best Local Similarity 88.7%; Pred. No. 1.7e-211;  
Matches 582; Conservative 0; Mismatches 1; Indels 73; Gaps 1;

QY 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQSSYGTGYGPTDVSYTOAQTAT 60  
DB 9 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQSSYGTGYGPTDVSYTOAQTAT 68  
QY 61 YGQTAATSYGQPTGTTTPTAPQAVSQPVQGYGTGAVDTTTAVTTTQASVAAQSAVGT 120  
DB 69 YGQTAATSYGQPTGTTTPTAPQAVSQPVQGYGTGAVDTTTAVTTTQASVAAQSAVGT 128  
QY 121 QPAPYAGQOPATAATPAPDGNKPTETSQPSSGTGYNQPSLGYGQSNYSYPOVPSYP 180  
DB 129 QPAPYAGQOPATAATPAPDGNKPTETSQPSSGTGYNQPSLGYGQSNYSYPOVPSYP 188  
QY 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240  
DB 189 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 248  
QY 241 PPTGTSYQAAPSOYSQOSSSYGOQSSYFRQDHPSSMGVYQGESGFGSGGENRSMGSDNR 300  
DB 249 PPTGTSYQAAPSOYSQOSSSYGOQSSYFRQDHPSSMGVYQGESGFGSGGENRSMGSDNR 273  
QY 301 GRRGGRGDRGMSRGRGGRGMSAGRGFPKPGPMDDEGDDLDLGPVPDDESDN 360  
DB 274 -----PMDDEGDDLDLGPVPDDESDN 295  
QY 361 SAIYVQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDEKTKPKGDAVSYEDP 420  
DB 296 SAIYVQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDEKTKPKGDAVSYEDP 355  
QY 421 PTKAAAVEMFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGP 480  
DB 356 PTKAAAVEMFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGP 415  
QY 481 GGPMDRGRGGRGGRGGRGPRGSRGNPSGGGVQHRADWQCPNPGCGNPNFARTTEC 540  
DB 416 GGPMDRGRGGRGGRGGRGPRGSRGNPSGGGVQHRADWQCPNPGCGNPNFARTTEC 475  
QY 541 NCKAPKPEBGLPPLPPPGDGRGGRGGRGGLMDRGPGGMFRGGRGGRGGRGGR 600  
DB 476 NCKAPKPEBGLPPLPPPGDGRGGRGGRGGLMDRGPGGMFRGGRGGRGGRGGR 535  
QY 601 GGRGMDRGRGGRGGRGGRGPGPLMEQMGRRGRGGRGPGKMDKGHRQRRDRPY 656  
DB 536 GGRGMDRGRGGRGGRGGRGPGPLMEQMGRRGRGGRGPGKMDKGHRQRRDRPY 591

RESULT 5  
US-09-949-016-10915

Sequence 10915, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FaSeq for Windows Version 4.0  
SEQ ID NO 10915  
LENGTH: 591  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-10915

Query Match 87.5%; Score 3178.5; DB 4; Length 591;  
Best Local Similarity 88.7%; Pred. No. 1.7e-211; Indels 73; Gaps 1;  
Matches 582; Conservative 0; Mismatches 1;

QY 1 MASTDVSYSQAQAAGGSAVTAQPTQGAQTTOAVGQSSYGTGQPTDVSYTOAQTAT 60  
DB 9 MASTDVSYSQAQAAGGSAVTAQPTQGAQTTOAVGQSSYGTGQPTDVSYTOAQTAT 68  
QY 61 YGQTATATSGQPTGTTTTPAQAQVSGVGTGAYDTTATVTTTQASVAAQSAVGT 120  
DB 69 YGQTATATSGQPTGTTTTPAQAQVSGVGTGAYDTTATVTTTQASVAAQSAVGT 128  
QY 121 QPAYPAVGGQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
DB 129 QPAYPAVGGQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 188  
QY 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYQGPSYQGOSSYQGOSSYQGOPTSY 240  
DB 189 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYQGPSYQGOSSYQGOSSYQGOPTSY 248  
QY 241 PPTQTSYQAAPSOYSSQSSSYGQSSFRQHPBSMGVYQGESGSGFSGPGENRMSGPDNR 300  
DB 249 PPTQTSYQAAPSOYSSQSSSYGQOR----- 273  
QY 301 GRGRGGRGDRGMRGGRGGRGMSAGRGFNKPCGPMDEGPDLDLGPVDPDESDN 360  
DB 274 -----PMDGPDLDLGPVDPDESDN 235  
QY 361 SAIVYGLANDSVTLDDLADPFKCGGVVKNKRTGQPMIHTYLDKRTGKPGDAVSYEDP 420  
DB 296 SAIVYGLANDSVTLDDLADPFKCGGVVKNKRTGQPMIHTYLDKRTGKPGDAVSYEDP 355  
QY 421 PTKAAVEMFDGDFQSKLVSLARKKPPMNSMRGGLPREGRGMPPLRGGPGGGGR 480  
DB 356 PTKAAVEMFDGDFQSKLVSLARKKPPMNSMRGGLPREGRGMPPLRGGPGGGGR 415  
QY 481 GGMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
DB 416 GGMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 475  
QY 541 NCKKAPPEGFLPPPPPGGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
DB 476 NCKKAPPEGFLPPPPPGGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 535  
QY 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
DB 536 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 591

RESULT 6  
US-08-437-027-21

Sequence 21, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landanyi, Marc  
APPLICANT: Gerald, William  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21  
CORRESPONDENCE ADDRESS: 21

ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,027  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 46416/JPM/CCA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-437-027-21

Query Match 39.1%; Score 1420; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 2.2e-90; Indels 0; Gaps 0;  
Matches 265; Conservative 0; Mismatches 0;

QY 1 MASTDVSYSQAQAAGGSAVTAQPTQGAQTTOAVGQSSYGTGQPTDVSYTOAQTAT 60  
DB 1 MASTDVSYSQAQAAGGSAVTAQPTQGAQTTOAVGQSSYGTGQPTDVSYTOAQTAT 60  
QY 61 YGQTATATSGQPTGTTTTPAQAQVSGVGTGAYDTTATVTTTQASVAAQSAVGT 120  
DB 61 YGQTATATSGQPTGTTTTPAQAQVSGVGTGAYDTTATVTTTQASVAAQSAVGT 120  
QY 121 QPAYPAVGGQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
DB 121 QPAYPAVGGQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
QY 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYQGPSYQGOSSYQGOSSYQGOPTSY 240  
DB 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYQGPSYQGOSSYQGOSSYQGOPTSY 240  
QY 241 PPTQTSYQAAPSOYSSQSSSYGQSSFRQHPBSMGVYQGESGSGFSGPGENRMSGPDNR 300  
DB 241 PPTQTSYQAAPSOYSSQSSSYGQSSFRQHPBSMGVYQGESGSGFSGPGENRMSGPDNR 300

RESULT 7  
US-08-437-027-20

Sequence 20, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landanyi, Marc  
APPLICANT: Gerald, William  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/437,027  
 FILING DATE:  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 46416/JPW/CCA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400  
 TELEFAX: 212-391-0525  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 365 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-437-027-20

Query Match 39.1%; Score 1420; DB 1; Length 365;

Best Local Similarity 100.0%; Pred. No. 2,2e-90;

Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDSTVSGAAGGYSAYTAQPTQGYAQTQAGYGGQSGYGTGPTDVSSTQAQTAT 60  
 DB 1 MASTDSTVSGAAGGYSAYTAQPTQGYAQTQAGYGGQSGYGTGPTDVSSTQAQTAT 60  
 QY 61 YGOTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSAVGT 120  
 DB 61 YGOTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSAVGT 120  
 QY 121 QPAPYPAAGQPPATAPRPDQNKPTETSPQSTGTGYNQPSLGGYGSNYSYQVPGSYP 180  
 DB 121 QPAPYPAAGQPPATAPRPDQNKPTETSPQSTGTGYNQPSLGGYGSNYSYQVPGSYP 180  
 QY 181 MOVVTAPEPSYPTSYSTQPTSYDQSSYSQONTTGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 DB 181 MOVVTAPEPSYPTSYSTQPTSYDQSSYSQONTTGGPSSYGGQSSYGGQSSYGGQPTSY 240  
 QY 241 PQQTGYSQAPSYGSSQSSSYGGQSS 265  
 DB 241 PQQTGYSQAPSYGSSQSSSYGGQSS 265

## RESULT 8

US-09-949-016-7825  
 Sequence 7825, Application US/09949016  
 Patent No. 6812339  
 GENERAL INFORMATION:  
 APPLICANT: VENTER, J. Craig et al.  
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 FILE REFERENCE: C1001307  
 CURRENT APPLICATION NUMBER: US/09/949,016  
 CURRENT FILING DATE: 2000-04-14  
 PRIOR APPLICATION NUMBER: 60/241,755  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/237,768  
 PRIOR FILING DATE: 2000-10-03  
 PRIOR APPLICATION NUMBER: 60/231,498  
 PRIOR FILING DATE: 2000-09-08  
 NUMBER OF SEQ ID NOS: 207012  
 SOFTWARE: FASTSEQ for Windows Version 4.0  
 SEQ ID NO 7825  
 LENGTH: 306  
 TYPE: PRT  
 ORGANISM: Human

## US-09-949-016-7825

Query Match 36.8%; Score 1337; DB 4; Length 306;  
 Best Local Similarity 78.3%; Pred. No. 9,6e-85;  
 Matches 246; Conservative 13; Mismatches 37; Indels 18; Gaps 2;

QY 339 PMDEGPDLDGPPVDDDESDNSAIYVGGINDSVTLDDLAFKQGVYMNKRTQPMI 398  
 DB 10 PMDEGPDLDGPPVDDDESDNSAIYVGGINDSVTLDDLAFKQGVYMNKRTQPMI 69  
 QY 399 HTYLDKTKGPKDATTATSYEDPTAXAAVEMFPGKDFQSGSKLTVSLARKKPPMNSRGL 458  
 DB 70 HTYLDKTKGPKDATTATSYEDPTAXAAVEMFPGKDFQSGSKLTVSLARKKPPMNSRGL 129  
 QY 459 PPRGGMPPPLNGRPGPGPGPGPMGMRGGRGDRGPPRGRGSRGNPSGGVYQHR 518  
 DB 130 PPRGGMPPPLNGRPGPGPGPGPMGMRGGRGDRGPPRGRGSRGNPSGGVYQHR 189  
 QY 519 AGDMQCPNCGGNQNTAMRTKCNQCKAPKREGLPPPPPPGDRGRGGRGGRGL 578  
 DB 190 AGDMQCPNCGGNQNTAMRTKCNQCKAPKREGLPPPPPPGDRGRGGRGGRGL 249  
 QY 579 MDRGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 638  
 DB 250 MDRGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 291  
 QY 639 PGMDKGEHRQERR 652  
 DB 292 PRTEPVTNKRKKR 305

## RESULT 9

US-09-538-092-1080  
 Sequence 1080, Application US/09538092  
 Patent No. 6753314  
 GENERAL INFORMATION:  
 APPLICANT: Glot, Loic  
 APPLICANT: Mansfield, Traci A.  
 TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 FILE REFERENCE: 15966-542  
 CURRENT APPLICATION NUMBER: US/09/538,092  
 CURRENT FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: 60/127,352  
 PRIOR FILING DATE: 1999-04-01  
 PRIOR APPLICATION NUMBER: 60/178,965  
 PRIOR FILING DATE: 2000-02-01  
 NUMBER OF SEQ ID NOS: 1387  
 SOFTWARE: CuratSeqFormatter Version 0.9  
 SEQ ID NO 1080  
 LENGTH: 526  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (0)..(0)  
 OTHER INFORMATION: Polypeptide Accession Number P35637  
 US-09-538-092-1080

Query Match 35.3%; Score 1282; DB 4; Length 526;  
 Best Local Similarity 45.0%; Pred. No. 1,1e-80;  
 Matches 301; Conservative 59; Mismatches 153; Indels 156; Gaps 30;

QY 1 MASTDSTVSGAAGGYSAYTAQPTQGYA-QTTQAGYGGQSGYGTGPTDVSSTQAQTAT 59  
 DB 1 MASTDSTVSGAAGGYSAYTAQPTQGYA-QTTQAGYGGQSGYGTGPTDVSSTQAQTAT 59  
 QY 60 TYGQTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSA 117  
 DB 49 TYGQTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSA 90  
 QY 118 YGOTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSA 177  
 DB 91 YGOTAVATSYGQPTGYTTPTAPQAVSQPVQYGTGAYDPTTAVTTTQASVAAQSA 128

QY 178 SYMOPVATAPSPPTSSSTOPTSYDQSSYQONTYQSPSSYQSSSYGQSSSYGQPP 237  
 Db 129 SYGQP-----SYG-----OQSYGQOQSTNPPQYIGQOQYNSSGCGGCG 172  
 QY 238 TSYPPQTSYQAPSOYSSQSSSYGQSSSPRODHPSSMGVYQSGSGSPGENRMSGP 237  
 Db 173 G-----GNYGQDQSSMSGSGSGGCGYQNDQSGCGSGGCGYQDQD-----GGRGSG- 222  
 QY 298 DNRGRGGGDR--GGMSSRGGRGGGGRGMSAG--ERGGRKGGGPPMDDEPDLDLGPV 353  
 Db 223 GGGGGGGGGRSSGGEPRGRGGGRGGRGMSGSDRGGRNKGFGPRDQSSRD-----S 277  
 QY 354 PDESDSAYVYQSSSYTLDLADPPKOCGVYKANKRFGQPMIHTYLDKRGKPKGDA 413  
 Db 278 EQDSDNNTTFVQGLGNVYIESVADYFKQIGIITKTKKQPMINITYTREGKLGKA 337  
 QY 414 TVSYEDPPTAAVWEPDGDPOGSKLVSLARKKPPMNSMRGGLPREGRGMPPLRG 473  
 Db 338 TVSFDPDPSAKAALDWFDKBFGNPIKVSPTATRAAFN--RGCGNRRGGRG-----RG 390  
 QY 474 PGPRGGGPGMGKMGSGGDRGCGPPPRGGRSGRNGPCCGVQHRADNOCPPPGCQNO 533  
 Db 391 PMGSGYGGG-----GSGGGRGGGPPSG-----GCGGQQRAGDMXCPNPTCENMN 437  
 QY 534 FAWRTCNOCAPKPEBFLPPPPPGDGRGGRGPGGMR--GGRGLMDRGPGGMPRG-- 590  
 Db 438 FSRNRCNOCAPKPEPDG-----PG-----GPGGSHMGYNGYDGRGRGGYDRG 484  
 QY 591 -GRGDRGGRGGR--GMDRGFGGGRGGRGPPPLMEOWGGRGGRGPGKMD--KGEH 647  
 Db 485 RGRGDRGGRGGRGGRGGRG-----GPGKMDSGR 517  
 QY 648 ROERDRPY 656  
 Db 518 RODRERPY 526

RESULT 10  
 US-08-343-443B-107  
 Sequence 107, Application US/08343443B

Patent No. 5968734  
 GENERAL INFORMATION:  
 APPLICANT: Aurias, Alain  
 APPLICANT: Delattre, Olivier  
 APPLICANT: Desmazes, Chantal  
 APPLICANT: Melot, Thomas  
 APPLICANT: Ploougaestel, Beatrice  
 APPLICANT: Thomas, Gilles  
 APPLICANT: Zucman, Jessica  
 TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
 TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
 TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
 TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
 TITLE OF INVENTION: TRANSLOCATIONS  
 NUMBER OF SEQUENCES: 129  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Weiser & Associates  
 STREET: 230 South Fifteenth Street  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19102  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: AEDIT 1.0 DOS text editor  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/343.443B  
 FILING DATE: 18-NOV-1994  
 CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/FR93/00494  
 FILING DATE: 19-MAY-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 92/06123  
 FILING DATE: 20-MAY-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Weiser, Gerard J.  
 REGISTRATION NUMBER: 19,763  
 REFERENCE/DOCKET NUMBER: 989.6121P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-875-8383  
 TELEFAX: 215-875-8394  
 INFORMATION FOR SEQ ID NO: 107:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 295 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-343-443B-107

Query Match 13.3%; Score 484; DB 2; Length 295;  
 Best Local Similarity 100.0%; Pred. No. 6,7e-26;  
 Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 PHSYPPQTSYQAPSOYSSQSSSYGQSSSPRODHPSSMGVYQSGSGSPGENRMSG 236  
 Db 1 PHSYPPQTSYQAPSOYSSQSSSYGQSSSPRODHPSSMGVYQSGSGSPGENRMSG 60  
 QY 297 PDNRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 325  
 Db 61 PDNRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 89

RESULT 11  
 US-09-919-039-324  
 Sequence 324, Application US/09919039

Patent No. 6727066  
 GENERAL INFORMATION:  
 APPLICANT: Kaser, Matthew R.  
 TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
 FILE REFERENCE: PA-0035 US  
 CURRENT APPLICATION NUMBER: US/09/919, 039  
 CURRENT FILING DATE: 2002-09-09  
 PRIOR APPLICATION NUMBER: 60/222, 113  
 PRIOR FILING DATE: 2000-07-28  
 NUMBER OF SEQ ID NOS: 401  
 SOFTWARE: PERL Program  
 SEQ ID NO 324  
 LENGTH: 462  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: Incyte ID No. 6727066 1813444CD1  
 US-09-919-039-324

Query Match 13.0%; Score 470.5; DB 4; Length 462;  
 Best Local Similarity 38.1%; Pred. No. 9,7e-25;  
 Matches 144; Conservative 25; Mismatches 92; Indels 101; Gaps 19;

QY 1 MASTDYSYTSQAAGSYSAVTAQPTQGYA-OTTAAYGQSSYGTGPTDVSYTAQTTA 59  
 Db 1 MASTDYSYTSQAAGSYSAVTAQPTQGYA-OTTAAYGQSSYGTGPTDVSYTAQTTA 48  
 QY 60 TGGQTAATSYGQPP-TGYTTPTAPQAYSQPVQYV-TGAYDTTATVTTTQASVAQSA 117  
 Db 49 GYGQSSY-SYSGSSQNTGVTGOSTP-----QYVSTGQYV-----SSGSSSS 90  
 QY 118 YGTQPAIYPAQGPATATPQDGNKPTETGQPOSTGYNOPSIGYQSNYSYQVPG 177  
 Db 91 YGQSSYSPYGOQPA-----PSSSTGSSYSSS---OSSSYGQPO-SG 128



```

Db 111 GPPGKMGHRGPDNGEKGEGTGAAGLKGNGLPNGENGAAPGMPGRAP----- 158
Qy 200 PTSYDSSSYQNTYQTPSSYGOSSSYGOSSSYGOQPTPTPTQTSYSAAPSYGOSS 259
Db 159 -----GERGPRGLPGAAGARGSDQPPPG-PPGTAGFPSPG-----AKG 204
Qy 260 SYGOSSFRDHPSSMVGQ--ESG--GFSG-----GEMRSMGPDNRG----- 301
Db 205 EVGPAGS-----PGSNAPGQGRGEPGQHAQAQPPGPGIINGSPGKEMGPAGIPGA 259
Qy 302 -----RGRGPDRCGMSRGGRGGRGMSAGRGGRGPNKPGPMDEGPD 346
Db 260 PGLMGARPPGPAAGANGAPGL-RGAGBPGKNGAKGPRGRGEAGIPIGVGAKGEDG 318
Qy 347 DLGPVDPDSDNSALTYVQGLNDSTLDDLADFFQCCGVVKNKRTGQPMHIYDKET 406
Db 319 KDSPGPGPANGLPGLAAGERGALGS----- 343
Qy 407 GKPKGATVSEYDPTAKAAVEMFDKDFQSKLVSLARKKPPMNSMRG--GLPPREG 464
Db 344 RGPAGPNGIPGKGPAGE-----RGAGPAGPAGAGPGRDGV 382
Qy 465 GMPPLRGPRGPRGPG--GPMGRWG--GRGDRGFPFPPRGRSG--NPSGG 512
Db 383 PGSPGMRGMPGSPGSGDKPKPGSGESGRPPG--PSGPRGQPGVMGFPBGKN 439
Qy 513 GNVQHRAGDMQCPNPGCGNPFAMRTCCNCKAPKPGFLPPPPP-----PGDGRGGR 568
Db 440 DGAPGKNGRGGPG--GPGPG--PPGKNGEYGPQGPPTGPGDKDGTGP 487
Qy 569 GGMRG-----GRGGLMDR-----GGPGMFRGGRGDRG-----GF 599
Db 488 RGPQGLQGLPTGTGPPENKPGEPKGBAAGAPAGCGKADGAPBERPPGLAAGPGL 547
Qy 600 RGRGMDRGPRGGRGPRGPG--PLMEOWGRRGRGPRG--KMDXGE 646
Db 548 RCGAG-PPGEGGKGAAGPPGPAAGTGLQGMFERGGLSGFPGKDXGE 598

```

# RESULT 14

US-09-500-811-21

Sequence 21, Application US/09500811

Patent No. 6323314

GENERAL INFORMATION:

APPLICANT: Ovist, Per

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/500,811

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Gogor18, Adda C

REGISTRATION NUMBER: 29,714

```

REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21
Query Match 10.4%; Score 376.5; DB 3; Length 1078;
Best Local Similarity 25.8%; Pred. No. 8e-18;
Matches 184; Conservative 34; Mismatches 235; Indels 259; Gaps 34;

```

```

Qy 42 GTTGPTDVSYDAQTATATYGTATATSYGQPTTPTTPTAQAISQPIQYGTGAYDTT 101
Db 39 GTSGRP-----GSPGSPGYGPPGER--GGAGSGPPG--PGALG----- 75
Qy 102 TATVTTQASVAAQSAVGTGPAYPAVGOQPAATAPTRPDGKPTETSGPQSGTGVN 161
Db 76 -----PSGPA-GKDGESGRPRPDRLGP--GPGIKGAPAGLP 110
Qy 162 SL-----GYQSNYSYPO-----VPGSY--PMQVTPAPSPPTSYSTQ 199
Db 111 GPPGKMGHRGPDNGEKGEGTGAAGLKGNGLPNGENGAAPGMPGRAP----- 158
Qy 200 PTSYDSSSYQNTYQTPSSYGOSSSYGOSSSYGOQPTPTPTQTSYSAAPSYGOSS 259
Db 159 -----GERGPRGLPGAAGARGSDQPPPG-PPGTAGFPSPG-----AKG 204
Qy 260 SYGOSSFRDHPSSMVGQ--ESG--GFSG-----GEMRSMGPDNRG----- 301
Db 205 EVGPAGS-----PGSNAPGQGRGEPGQHAQAQPPGPGIINGSPGKEMGPAGIPGA 259
Qy 302 -----RGRGPDRCGMSRGGRGGRGMSAGRGGRGPNKPGPMDEGPD 346
Db 260 PGLMGARPPGPAAGANGAPGL-RGAGBPGKNGAKGPRGRGEAGIPIGVGAKGEDG 318
Qy 347 DLGPVDPDSDNSALTYVQGLNDSTLDDLADFFQCCGVVKNKRTGQPMHIYDKET 406
Db 319 KDSPGPGPANGLPGLAAGERGALGS----- 343
Qy 407 GKPKGATVSEYDPTAKAAVEMFDKDFQSKLVSLARKKPPMNSMRG--GLPPREG 464
Db 344 RGPAGPNGIPGKGPAGE-----RGAGPAGPAGAGPGRDGV 382
Qy 465 GMPPLRGPRGPRGPG--GPMGRWG--GRGDRGFPFPPRGRSG--NPSGG 512
Db 383 PGSPGMRGMPGSPGSGDKPKPGSGESGRPPG--PSGPRGQPGVMGFPBGKN 439
Qy 513 GNVQHRAGDMQCPNPGCGNPFAMRTCCNCKAPKPGFLPPPPP-----PGDGRGGR 568
Db 440 DGAPGKNGRGGPG--GPGPG--PPGKNGEYGPQGPPTGPGDKDGTGP 487
Qy 569 GGMRG-----GRGGLMDR-----GGPGMFRGGRGDRG-----GF 599
Db 488 RGPQGLQGLPTGTGPPENKPGEPKGBAAGAPAGCGKADGAPBERPPGLAAGPGL 547
Qy 600 RGRGMDRGPRGGRGPRGPG--PLMEOWGRRGRGPRG--KMDXGE 646
Db 548 RCGAG-PPGEGGKGAAGPPGPAAGTGLQGMFERGGLSGFPGKDXGE 598

```

# RESULT 15

US-09-570-573-21

Sequence 21, Application US/09570573

Patent No. 5342361  
 GENERAL INFORMATION:  
 APPLICANT: Ovislc, Per  
 APPLICANT: Bonde, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
 TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
 NUMBER OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dardy & Dardy PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/570,573  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogorile, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236687  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1078 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (III)  
 IS-09-570-573-21

Query Match	10.4%	Score 376.5	DB 3	Length 1078
Best Local Similarity	25.8%	Pred. No. 8e-18		
Matches 184	Conservative 34	Mismatches 235	Indels 259	Gaps 34
QY	42	GYVGQPTDVSYTAQTTATVGTAVTASVYGQPTPTTPTTAPOAVSQPVGQVGTGAVDTT	101	
DB	39	GTSHHP-----GSPGSPVQGPPEP---GQAPFSQPPG---PPAIG-----	75	
QY	102	TATVTTTQASVTAOSAVGTOPAVPAYVGGQAPATAETPRPDGKPKPTETSOPOSTGYNQP	161	
DB	76	-----PSGPA-GKDGSSGRPRGDRGLP---GPPQIKGPAPIG	110	
QY	162	SL-----GYGGSNYSYPQ-----VPSGY---PMQPTAPPTSPPTSYSSTQ	199	
DB	111	GFPCKMKHRRGDGNGEKETGATGALCKENGLPGENGAPGPMGPRGAP-----	158	
QY	200	PTSYDQSSVSPQONTYYGQPSYVGQSSVYGQSSVYGQPPQPTSYVPPQGSYSQA.PSOYSSQSS	259	
DB	159	-----GERGPRGLPGAAGAKRNDGARSDDGPPPG--PPGTAAGPQSPG---AKG	204	
QY	260	SYGQSSSFRODHPSPSMGVYQO--ESG--GFGS-----PEENRSMSPDNRG--	301	
DB	205	EYVPAQS-----PSNGAPQGRGEGPCGGHAGAGACPPPPQTLNNGSPGKKGEMGPAIGTGA	259	
QY	302	-----RGRGGRFDRGMSRGRGGRGGMGASGERGFFNKPQGPMDCEGDL	346	

Db 260 PGLMGARGP PGCPAGANGAPGL -RGAGAGEPGKNGAAGEPGCPBERGEAGI PGVPGAAGEDG 318  
QY 347 DLGPEVDEDESDNSAIVYQGLNDSVTLDDLADLFFKQCGVVMKMKRGTQPMHIYLDKET 406  
Db 319 KDGSGLDGGANGLPALAGERGLSS----- 343  
QY 407 GKPKGDATVSYEDPEPTAAVAWEVDFDKDFQSGSLVSLARKKP MNMSMG -GLPPREG 464  
Db 344 RGPAGPNCI PGEBKPAGE----- RGA PGAPGPGAAGEPGRDGV 382  
QY 465 GMPPLRLRGPGPGPGPG----- GPMGRNG -GRGDRCGFPPRPGPSRG----- NPSGG 512  
Db 383 PGCPGMRGMPGSPGPGSGDGKPGPGSGSGSRPPPG---PSGPGQGVVMGFPGPKGN 439  
QY 513 GNVHRADDMQCPNPGCGCNQNPAMVTECNCQCAPRPEGLPPPPPP-----PGDRRRGGR 568  
Db 440 DGAPGPKNGERGPG -GPGPG----- PPKNGEYVGQGP GPGPTGPGDGKDDTGP 487  
QY 569 GGMG-----GRGGLMDR-----GGPGMFRGGGRDGR-----GP 539  
Db 488 RGPQOLQGLPGTGP PGGENEKPGEPCPGKEAGAPGAPGKGDA GAGBERGPGGLAGAPGL 547  
QY 600 RGRGMDRCGRFGGGRGPGPGPG-----PLMEGARGRRGGRGPG -KMDKGE 646  
Db 548 RGGAG -PPGPGCGKGAAGPPPGPGAAGTGTGLOGMGRBERGLGSPGPKDKGE 598

Search completed: February 18, 2005, 15:13:48  
Job time : 50 secs



**This Page Blank (uspto)**

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 18, 2005, 15:12:20 ; Search time 135 Seconds  
(without alignments)  
1590.149 Million cell updates/sec

Title: US-10-791-017A-2  
Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAGGYSYA.....GGGKNDKGRHRRDRPY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues  
Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3633	100.0	656	US-10-791-017A-2	Sequence 2, Appl1
2	3277	90.2	600	US-10-094-749-2713	Sequence 2713, Ap
3	1281.5	35.3	525	US-10-755-889-650	Sequence 650, App
4	799.5	22.0	260	US-10-408-765A-116	Sequence 116, App
5	623	17.1	156	US-09-925-301-1511	Sequence 1511, Ap
6	475	13.1	83	US-09-864-761-42397	Sequence 42397, A
7	470.5	13.0	462	US-09-919-039-3324	Sequence 324, App
8	423	11.6	467	US-10-437-963-191926	Sequence 191926,
9	420	11.6	395	US-10-435-114-69744	Sequence 69744, A
10	400.5	11.0	1466	US-09-918-715-226	Sequence 226, App
11	400.5	11.0	1466	US-10-177-293-68	Sequence 68, Appl
12	400.5	11.0	1466	US-10-301-822-33	Sequence 33, Appl
13	400.5	11.0	1466	US-10-257-021-72	Sequence 72, Appl

14	400.5	11.0	1466	US-10-357-851-3	Sequence 3, Appl1
15	400.5	11.0	1466	US-10-358-024-3	Sequence 3, Appl1
16	400.5	11.0	1466	US-10-734-564-103	Sequence 103, App
17	388.5	10.7	1466	US-10-402-089-4	Sequence 4, Appl1
18	388.5	10.7	1466	US-10-402-089-6	Sequence 6, Appl1
19	388.5	10.7	1466	US-10-402-072A-4	Sequence 4, Appl1
20	388.5	10.7	1466	US-10-402-072A-6	Sequence 6, Appl1
21	384	10.6	1466	US-10-402-089-12	Sequence 12, Appl
22	384	10.6	1466	US-10-402-089-12	Sequence 12, Appl
23	376.5	10.4	1078	US-10-058-124-21	Sequence 21, Appl
24	348.5	9.6	1449	US-10-402-089-8	Sequence 8, Appl1
25	348.5	9.6	1449	US-10-402-072A-8	Sequence 8, Appl1
26	348	9.6	1496	US-10-177-293-70	Sequence 70, Appl
27	348	9.6	1496	US-10-301-822-35	Sequence 35, Appl
28	348	9.6	1496	US-10-236-031B-74	Sequence 74, Appl
29	348	9.6	1496	US-10-468-091-22	Sequence 22, Appl
30	348	9.6	1496	US-10-468-091-22	Sequence 22, Appl
31	346.5	9.5	1496	US-10-788-792-248	Sequence 248, App
32	346.5	9.5	1497	US-10-437-963-164510	Sequence 164510,
33	346	9.5	1463	US-10-468-091-23	Sequence 23, Appl
34	346	9.5	1463	US-10-402-089-2	Sequence 2, Appl1
35	344	9.5	1463	US-10-402-072A-2	Sequence 2, Appl1
36	337.5	9.3	1461	US-10-468-091-25	Sequence 25, Appl
37	335.5	9.2	1418	US-10-468-091-25	Sequence 25, Appl
38	335.5	9.2	1418	US-10-058-124-20	Sequence 20, Appl
39	334	9.2	1739	US-10-468-091-5	Sequence 5, Appl1
40	333.5	9.2	714	US-09-795-061-2	Sequence 2, Appl1
41	333.5	9.2	714	US-09-861-597-10	Sequence 10, Appl
42	333	9.2	854	US-10-887-100-10	Sequence 10, Appl
43	333	9.2	854	US-10-488-056-47	Sequence 47, Appl
44	333	9.2	1464	US-09-918-715-261	Sequence 261, App
45	333	9.2	1464	US-10-060-035-159	Sequence 159, App
			1464	US-10-171-311-36	Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-10-791-017A-2  
; Sequence 2, Application US/10791017A  
; Publication No. US20040197827A1  
; GENERAL INFORMATION:  
; APPLICANT: JENAPARM GmbH & Co. KG  
; TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances  
; FILE REFERENCE: Pat 3684/11  
; CURRENT APPLICATION NUMBER: US/10/791, 017A  
; CURRENT FILING DATE: 2004-03-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 656  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-791-017A-2

Query Match 100.0%; Score 3633; DB 16; Length 656;  
Best Local Similarity 100.0%; Pred. No. 4.7e-209;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MASTDSTYSQAAGGYSYAQTPTQAGYAGGOSYGYGPTDVSYQAQTAT 60  
DB 1 MASTDSTYSQAAGGYSYAQTPTQAGYAGGOSYGYGPTDVSYQAQTAT 60  
QY 61 YGOTAVATSYGAPPTGTTPTTAPOAVSOPVQGYGTGAYDTTATTTTQAASAYAGT 120  
DB 61 YGOTAVATSYGAPPTGTTPTTAPOAVSOPVQGYGTGAYDTTATTTTQAASAYAGT 120  
QY 121 QPAYVAYGQPAATPATRPODGKPKETISQPOSSTGYNQPSIGYQSNVSYQVPGSY 180  
DB 121 QPAYVAYGQPAATPATRPODGKPKETISQPOSSTGYNQPSIGYQSNVSYQVPGSY 180  
QY 181 MGVTPAPSPYSPSTGPTSTGSSYQSNVYQPSIGYQSNVSYQVPGSY 240  
DB 181 MGVTPAPSPYSPSTGPTSTGSSYQSNVYQPSIGYQSNVSYQVPGSY 240

```

Db 181 MGPVTAPEPSYPTSTSTPTSTVDSSYSQONTYGPSSYGGQSSYGGQPTSTY 240
Qy 241 PPTGTSYGAAPSOYSGQSSSYGQSSFRQDHPSSMGVYGGESGFGSGPGENRSMSPDNR 300
Db 241 PPTGTSYGAAPSOYSGQSSSYGQSSFRQDHPSSMGVYGGESGFGSGPGENRSMSPDNR 300
Qy 301 GRGRGFGDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
Db 301 GRGRGFGDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
Qy 361 SAIVYGLNDSVTLDDLADFPKCGVYKNNKRTGQPMIHIYLDKXETGKPGDATVSIEDP 420
Db 361 SAIVYGLNDSVTLDDLADFPKCGVYKNNKRTGQPMIHIYLDKXETGKPGDATVSIEDP 420
Qy 421 PTKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRRGRGMPPLRGPGGPGGP 480
Db 421 PTKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRRGRGMPPLRGPGGPGGP 480
Qy 481 GGRMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
Db 481 GGRMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
Qy 541 NCKAPKPEGFLPPPPPPGDRGGRGGRGGRGGRGGLMDRGPGGMFRGGRGDRGGR 600
Db 541 NCKAPKPEGFLPPPPPPGDRGGRGGRGGRGGRGGLMDRGPGGMFRGGRGDRGGR 600
Qy 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656
Db 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 2

```

US-10-094-749-2713
; Sequence 2713, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKA, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORI
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2713
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2713

```

```

Query Match 90.2%; Score 3277; DB 15; Length 600;
Best Local Similarity 91.3%; Pred. No. 7.9e-188;
Matches 599; Conservative 0; Mismatches 1; Indels 56; Gaps 1;

```

```

Qy 1 MASTDSTYSQAAPSOYSAATTAOPTGYAOTTAOYAGQSSYGVYGGQPTDVSYTOAQTAT 60
Db 1 MASTDSTYSQAAPSOYSAATTAOPTGYAOTTAOYAGQSSYGVYGGQPTDVSYTOAQTAT 60
Qy 61 YGOTAAVATSYGAPPTGYTTPTTAPOAVSPVQGYGTGAYVTTTATVTTTAAVAAQAGYT 120
Db 61 YGOTAAVATSYGAPPTGYTTPTTAPOAVSPVQGYGTGAYVTTTATVTTTAAVAAQAGYT 120
Qy 121 GPAPYAGQOAPATAPTRPDQGNKPTETSQPOSSTGCVNQPSLGYQSNYSYTPQVGSYTP 180
Db 121 GPAPYAGQOAPATATATATATATATATATATATATATATATATATATATATATATAT 180
Qy 181 MGPVTAPEPSYPTSTSTPTSTVDSSYSQONTYGPSSYGGQSSYGGQPTSTY 240
Db 181 MGPVTAPEPSYPTSTSTPTSTVDSSYSQONTYGPSSYGGQSSYGGQPTSTY 240
Qy 241 PPTGTSYGAAPSOYSGQSSSYGQSSFRQDHPSSMGVYGGESGFGSGPGENRSMSPDNR 300
Db 241 PPTGTSYGAAPSOYSGQSSSYGQSSFRQDHPSSMGVYGGESGFGSGPGENRSMSPDNR 300
Qy 301 GRGRGFGDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
Db 301 GRGRGFGDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
Qy 361 SAIVYGLNDSVTLDDLADFPKCGVYKNNKRTGQPMIHIYLDKXETGKPGDATVSIEDP 420
Db 361 SAIVYGLNDSVTLDDLADFPKCGVYKNNKRTGQPMIHIYLDKXETGKPGDATVSIEDP 420
Qy 421 PTKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRRGRGMPPLRGPGGPGGP 480
Db 421 PTKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRRGRGMPPLRGPGGPGGP 480
Qy 481 GGRMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
Db 481 GGRMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
Qy 541 NCKAPKPEGFLPPPPPPGDRGGRGGRGGRGGRGGLMDRGPGGMFRGGRGDRGGR 600
Db 541 NCKAPKPEGFLPPPPPPGDRGGRGGRGGRGGRGGLMDRGPGGMFRGGRGDRGGR 600
Qy 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656
Db 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 3

```

US-10-755-889-650
; Sequence 650, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 650
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-650

```

```

Query Match 35.3%; Score 1281.5; DB 16; Length 525;
Best Local Similarity 44.8%; Pred. No. 1.1e-68;
Matches 299; Conservative 59; Mismatches 155; Indels 155; Gaps 29;

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 18, 2005, 14:57:35 ; Search time 44 Seconds

(without alignments)  
1434.504 Million cell updates/sec

Title: US-10-791-017a-2

Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAGGYSYA.....GGGKMDKXGHRQDRDPY 656Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 9621673 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 79:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3633	100.0	656	1 A49358	RNA-binding protei
2	3571.5	98.3	655	1 A55726	RNA-binding protei
3	1282	35.3	526	1 S33799	RNA-binding protei
4	1270.5	35.0	528	2 G02127	fus-like protein -
5	1003.5	27.6	589	2 S71954	RNA/sgDNA-binding
6	724	19.9	404	2 S54729	RNA-binding protei
7	554	15.2	545	2 T15667	hypothetical prote
8	484	13.3	295	4 S36174	RNA binding protei
9	472.5	13.0	462	4 S33798	FUS/CHOP mutant fu
10	400.5	11.0	1466	1 CGHUTL	collagen alpha 1(I
11	376.5	10.4	886	2 T50694	collagen alpha 1(I
12	374	10.3	1464	2 S59856	collagen alpha 1(I
13	370.5	10.2	211	2 C96539	hypothetical prote
14	368	10.1	1049	1 CGB07S	collagen alpha 1(I
15	359	9.9	1838	1 CGH01V	collagen alpha 1(I
16	348	9.6	1496	1 CGH02V	collagen alpha 2(V
17	346.5	9.5	1497	2 T49607	procollagen type V
18	344	9.5	1453	2 S21626	collagen alpha 1(I
19	344	9.5	1843	2 S18803	collagen alpha 1(I
20	343.5	9.5	1880	2 T18531	collagen - medicina
21	342.5	9.4	1414	1 S23809	collagen alpha 2(I
22	340.5	9.4	848	2 S02262	glutinin high mole
23	339.5	9.3	1024	2 S18251	collagen alpha 1(I
24	339.5	9.3	1482	2 A40333	collagen alpha 1(I
25	338.5	9.3	1486	1 B40333	collagen alpha 1(I
26	336.5	9.3	1418	2 T45467	collagen alpha 1(I
27	336	9.2	838	1 EEWTHM	glutinin, high mol
28	336	9.2	1752	2 A45407	collagen alpha 3(I
29	335.5	9.2	1487	1 CGH06C	collagen alpha 1(I

30	333	9.2	1464	1 CGH01S	collagen alpha 1(I
31	330	9.1	1806	1 CGH01E	collagen alpha 1(X
32	328	9.0	1027	2 S28774	collagen alpha cha
33	328	9.0	2944	2 A54849	collagen alpha 1(V
34	324	8.9	671	1 CGRT1S	collagen alpha 1(I
35	324	8.9	1585	2 T31611	hypothetical prote
36	322.5	8.9	1747	2 A54121	collagen alpha-4 c
37	321.5	8.9	3198	2 A43426	collagen alpha 2 f
38	320.5	8.8	1042	1 CGCH1S	collagen alpha 1(I
39	318	8.8	1142	2 T00022	B120 protein - hum
40	317	8.7	1419	2 A41182	collagen alpha 1(I
41	317	8.7	1487	2 B41182	collagen alpha 1(I
42	316.5	8.7	1670	1 CGH13B	collagen alpha 3(I
43	316	8.7	753	2 JC2099	glutinin, high mol
44	316	8.7	921	2 S40495	collagen alpha 1(I
45	315	8.7	921	2 S42617	collagen alpha 1(I

## ALIGNMENTS

RESULT 1  
A49358  
RNA-binding protein EMS - human  
C/Species: Homo sapiens (man)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: A49358; S28257  
R/Plougaestel, B.; Zucman, J.; Peter, M.; Thomas, G.; Delattre, O.  
Genomics 18, 609-615, 1993  
A/Title: Genomic structure of the EMS gene and its relationship to EMSR1, a site of tumor  
A/Reference number: A49358; MIM:9410360; PMID:8307570  
A/Accession: A49358  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-656 <RES>  
A/Cross-references: UNIPROT:Q01844; EMBL:X72990; NID:9485838; PIR:CAA51489.1; PID:982565  
R/Delattre, O.; Zucman, J.; Plougaestel, B.; Desmarte, C.; Melot, T.; Peter, M.; Kovar, H.,  
Nature 359, 162-165, 1992  
A/Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in  
A/Reference number: S28257; MIM:9236239; PMID:1522903  
A/Accession: S28257  
A/Molecule type: mRNA  
A/Residues: 1-656 <DEL>  
A/Cross-references: EMBL:X66899; NID:9547565; PIR:CAA47350.1; PID:931280  
C/Genetics:  
A/Genes: GDB:EMSR1  
A/Cross-references: GDB:135984; OMIM:133450  
A/Map position: 22q12.1-22q12.1  
A/Intons: 5/1; 17/2; 34/3; 76/1; 138/2; 265/1; 325/2; 338/1; 349/1; 388/3; 432/1.  
A/Note: EMSR1 region is exons 7-10 of this gene, called EMS in reference A49358; this reg  
C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology  
C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
F/362-437/Domain: ribonucleoprotein repeat homology <RMR>

Query Match 100.0%; Score 3633; DB 1; Length 656;  
Best Local Similarity 100.0%; Pred. No. 3.9e-168;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASTDSTYSQAAGGYSYAQTQAAGGYSYGTPDVSYQAQTAT	60
DB	1	MASTDSTYSQAAGGYSYAQTQAAGGYSYGTPDVSYQAQTAT	60
QY	61	YGGTAATATSYGPGTGTTPAPQAYSQPVGSGTGAADTTATVTTTQAASVAAGSAYGT	120
DB	61	YGGTAATATSYGPGTGTTPAPQAYSQPVGSGTGAADTTATVTTTQAASVAAGSAYGT	120
QY	121	QAPYPAVGGQPAATATPQDGNKPTETQPOSGTSGYNQPSLGGQSNVSYQVGVGSP	180
DB	121	QAPYPAVGGQPAATATPQDGNKPTETQPOSGTSGYNQPSLGGQSNVSYQVGVGSP	180
QY	181	MCQVTPAPSPYPTSYSTQPTSYDQSSYSGQNTYGGQSSYGGQSSYGGQSPPTSY	240
DB	181	MCQVTPAPSPYPTSYSTQPTSYDQSSYSGQNTYGGQSSYGGQSSYGGQSPPTSY	240

QY 241 PPTGTSYSAAPSOYSSQSSSYGQSSSFRDHPSSMGVYQGESGFGSGPGENRSMSPDNR 300  
 DB 241 PPTGTSYSAAPSOYSSQSSSYGQSSSFRDHPSSMGVYQGESGFGSGPGENRSMSPDNR 300  
 QY 301 GRGGGFDGRGMSRGGRGGGMSAGERGFNFKPGPMDEGPDLDLGPVDPDESDN 360  
 DB 301 GRGGGFDGRGMSRGGRGGGMSAGERGFNFKPGPMDEGPDLDLGPVDPDESDN 360  
 QY 361 SATYVGGLNDSTLDDLDLADFFKQCGVYKMKRTGQPMHITLIDKGTGKPGDATTSEDP 420  
 DB 361 SATYVGGLNDSTLDDLDLADFFKQCGVYKMKRTGQPMHITLIDKGTGKPGDATTSEDP 420  
 QY 421 PTKAAVWFEDGDFGSKLKVSLARKKPPMNSMRGLPPREGRGMPPLRGSGPGG 480  
 DB 421 PTKAAVWFEDGDFGSKLKVSLARKKPPMNSMRGLPPREGRGMPPLRGSGPGG 480  
 QY 481 GGPMMGRGGRGGRGGRGPPRPGSRGNPSGGGVQHRAGDMQCPNPGCGNPFARTEC 540  
 DB 481 GGPMMGRGGRGGRGGRGPPRPGSRGNPSGGGVQHRAGDMQCPNPGCGNPFARTEC 540  
 QY 541 NOCKAPKPEGFLPPPPPPGDRGGRGPGGMRGGGLMDRGPGGMFRGGRGDRGGR 600  
 DB 541 NOCKAPKPEGFLPPPPPPGDRGGRGPGGMRGGGLMDRGPGGMFRGGRGDRGGR 600  
 QY 601 GGRGMDRGFGGGRGGRGPGGPPGLMEQMGRRGGRGPGKMDKGRHROBRDRPY 656  
 DB 601 GGRGMDRGFGGGRGGRGPGGPPGLMEQMGRRGGRGPGKMDKGRHROBRDRPY 656

## RESULT 2

A55726  
 RNA-binding protein Ews - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: A55726; S45007  
 R/Plouffe, B.; Mactel, M.G.; Thomas, G.; Delattre, O.  
 Genomics 23, 278-281, 1994  
 A/Title: Cloning and chromosome localization of the mouse Ews gene.  
 A/Reference number: A55726; MUID:95130099; PMID:7823090  
 A/Accession: A55726  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-655 <PRO>  
 A/Cross-references: UNIPROT:Q61545; GB:X79233; NID:g488512; PID:g488513  
 A/Note: authors translated the codon TCA for residue 116 as Thr, GCC for residue 123 as C/GeneCis:  
 A/Gene: Ews  
 C/Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology  
 C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
 F/361-436/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 98.3%; Score 3571.5; DB 1; Length 655;  
 Best Local Similarity 98.0%; Pred. No. 3.6e-165;  
 Matches 643; Conservative 8; Mismatches 4; Indels 1; Gaps 1;  
 QY 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSYGYGPTDVSYTAQTAT 60  
 DB 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSYGYGPTDVSYTAQTAT 60  
 QY 61 YGOTAAVATSYGPPGTGTTPTAQAQVQVGYGTGATDTTATVTTTQAASAAAGAYGT 120  
 DB 61 YGOTAAVATSYGPPGTGTTPTAQAQVQVGYGTGATDTTATVTTTQAASAAAGAYGT 120  
 QY 121 QPAYPAVGGQPAATPAPDGNKPTETSGQSSGTGYNPSLGYGQSNYSYQVPGSYR 180  
 DB 121 QPAYPAVGGQPAATPAPDGNKPTETSGQSSGTGYNPSLGYGQSNYSYQVPGSYR 180  
 QY 181 MOGVNTPSPSPPTSYSTOFTSYDQSSYGOONTYGGPSSYGOQSSYGOQSPPTSY 240  
 DB 181 MOGVNTPSPSPPTSYSTOFTSYDQSSYGOONTYGGPSSYGOQSSYGOQSPPTSY 240  
 QY 241 PPTGTSYSAAPSOYSSQSSSYGQSSSFRDHPSSMGVYQGESGFGSGPGENRSMSPDNR 300

DB 241 PPTGTSYSAAPSOYSSQSSSYGQSSSFRDHPSSMGVYQGESGFGSGPGENRSMSPDNR 300  
 QY 301 GRGGGFDGRGMSRGGRGGGMSAGERGFNFKPGPMDEGPDLDLGPVDPDESDN 360  
 DB 301 GRGGGFDGRGMSRGGRGGGMSAGERGFNFKPGPMDEGPDLDLGPVDPDESDN 360  
 QY 361 SATYVGGLNDSTLDDLDLADFFKQCGVYKMKRTGQPMHITLIDKGTGKPGDATTSEDP 420  
 DB 361 SATYVGGLNDSTLDDLDLADFFKQCGVYKMKRTGQPMHITLIDKGTGKPGDATTSEDP 420  
 QY 421 PTKAAVWFEDGDFGSKLKVSLARKKPPMNSMRGLPPREGRGMPPLRGSGPGG 480  
 DB 421 PTKAAVWFEDGDFGSKLKVSLARKKPPMNSMRGLPPREGRGMPPLRGSGPGG 480  
 QY 481 GGPMMGRGGRGGRGGRGPPRPGSRGNPSGGGVQHRAGDMQCPNPGCGNPFARTEC 540  
 DB 481 GGPMMGRGGRGGRGGRGPPRPGSRGNPSGGGVQHRAGDMQCPNPGCGNPFARTEC 540  
 QY 541 NOCKAPKPEGFLPPPPPPGDRGGRGPGGMRGGGLMDRGPGGMFRGGRGDRGGR 600  
 DB 541 NOCKAPKPEGFLPPPPPPGDRGGRGPGGMRGGGLMDRGPGGMFRGGRGDRGGR 600  
 QY 601 GGRGMDRGFGGGRGGRGPGGPPGLMEQMGRRGGRGPGKMDKGRHROBRDRPY 656  
 DB 601 GGRGMDRGFGGGRGGRGPGGPPGLMEQMGRRGGRGPGKMDKGRHROBRDRPY 656

## RESULT 3

S33799  
 RNA-binding protein FUS, nuclear - human  
 N/Alternate names: RNA-binding protein TLS  
 C/Species: Homo sapiens (man)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: S33799; S36157  
 R/Crozat, A.; Aman, P.; Mandahl, N.; Ron, D.  
 Nature 363, 640-644, 1993  
 A/Title: Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma.  
 A/Reference number: S33799; MUID:93288139; PMID:8510758  
 A/Accession: S33799  
 A/Molecule type: mRNA  
 A/Residues: 1-526 <CRO>  
 A/Cross-references: UNIPROT:P35637; GB:S62140; NID:g386156; PID:AAB27102.1; PID:g386157  
 A/Experimental source: liposarcoma  
 R/Rabbit, T.H.; Porter, A.; Larson, R.; Nathan, P.  
 Nature Genet. 4, 175-180, 1993  
 A/Title: Fusion of the dominant negative transcription regulator CHOP with a novel gene  
 A/Reference number: S36157; MUID:93350637; PMID:7503811  
 A/Accession: S36157  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-63, 'S', 66-526 <RAB>  
 A/Cross-references: EMBL:X71428; NID:g393415; PID:CAA50559.1; PID:g4210363  
 A/Experimental source: liposarcoma  
 C/GeneCis:  
 A/Gene: GDB:FUS  
 A/Cross-references: GDB:136048; OMIM:137070  
 A/Map position: 16p11.2-16p11.2  
 C/Function:  
 A/Description: RNA binding; probable plays a role in transcriptional regulation  
 C/Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology  
 C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
 F/286-361/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 35.3%; Score 1282; DB 1; Length 526;  
 Best Local Similarity 45.0%; Pred. No. 4.7e-55;  
 Matches 301; Conservative 59; Mismatches 153; Indels 156; Gaps 30;  
 QY 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSYGYGPTDVSYTAQTAT 60  
 DB 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSYGYGPTDVSYTAQTAT 60  
 QY 60 TVGOTAAVATSYGPP-TGTTPTAQAQVQVGYGTGATDTTATVTTTQAASAAAGAYGT 117

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 14:55:56 ; Search time 177 Seconds

(without alignments)  
1897.875 Million cell updates/sec

Title: US-10-791-017a-2

Sequence: 1 MASTDYSTYSQAAAQGGYSA.....GGFGKMDKGRHQRDRPRY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_tramb1.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	656	1	EMS_HUMAN
2	3618.5	99.6	655	2	Q96FE8
3	3605.5	99.2	661	2	Q96MX4
4	3579	98.5	656	2	Q6NVA3
5	3571.5	98.3	655	1	EMS_MOUSE
6	3564.5	98.1	655	2	Q9D3P0
7	3277	90.2	600	2	Q96MNA
8	2601	71.6	674	2	Q6F3N0
9	2586.5	71.2	673	2	Q6DJQ4
10	1885.5	51.9	624	2	Q6NIX1
11	1874	51.6	633	2	Q9CRS5
12	1865	51.3	623	2	Q7SZR5
13	1755	48.3	354	2	Q9BWA2
14	1579	43.1	578	2	Q80J33
15	1566	43.5	575	2	Q6NNA7
16	1429	39.3	476	2	Q9BZD1
17	1332	36.4	504	2	Q6J4Y8
18	1332	36.4	536	2	Q7ZXQ2
19	1301	35.8	512	1	FUS_BOVIN
20	1290	35.5	518	1	FUS_MOUSE
21	1290	35.5	536	2	Q6DKB4
22	1289.5	35.5	517	2	Q6CFR9
23	1282	35.3	526	1	FUS_HUMAN
24	1277	35.2	526	2	Q8TBK3
25	1270.5	35.0	528	2	Q13344
26	1016.5	28.0	527	2	Q8BQ46
27	1009	27.8	532	1	R856_HUMAN
28	910	25.0	337	2	Q95KX8
29	907	25.0	337	2	Q86X94
30	904.5	24.9	501	2	Q6GDI1
31	900.5	24.8	482	2	Q6DCR4

32	884.5	24.3	475	2	Q66J39	Q66J39 xenopus lae
33	782	21.5	410	2	Q7ZUE3	Q7ZUE3 brachydanio
34	717.5	19.7	399	1	CAZ_DROME	Q27294 dirosophila
35	690	19.0	442	2	Q7PUK2	Q7PUK2 anopheles g
36	625.5	17.2	189	2	Q7IE78	Q7IE78 homo sapien
37	621	17.1	280	2	Q91VQ2	Q91VQ2 mus musculu
38	583.5	16.1	448	2	Q18265	Q18265 caenorhabdi
39	452.5	12.5	265	2	Q6DDA7	Q6DDA7 xenopus tro
40	423	11.6	414	2	Q69TN3	Q69TN3 oryza sativ
41	420.5	11.6	101	2	Q98UE8	Q98UE8 xenopus lae
42	414	11.4	394	2	Q6YS82	Q6YS82 oryza sativ
43	400.5	11.0	1163	2	Q8N6U4	Q8N6U4 homo sapien
44	400.5	11.0	1466	2	CA13_HUMAN	P02461 homo sapien
45	400	11.0	113	2	Q8BQ4	Q8BQ4 mus musculu

## ALIGNMENTS

RESULT 1	EMS_HUMAN	STANDARD;	PRT;	656 AA.
AC	001844; Q92635;			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	RNA-binding protein EMS (EWS oncogene) (Ewing sarcoma breakpoint region 1 protein).			
DE	Name=EWSR1; Synonyms=EWS;			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSTB-Petal brain;			
RX	MEDLINE=92396239; PubMed=1522903; DOI=10.1038/359162a0;			
RA	Delattre O., Zucman J., Plougaestel B., Desmaziere C., Melot T., Peter M., Kovar H., Joubert I., de Jong P., Rouleau G., Aurias A., Thomas G.;			
RT	"Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.";			
RT	Nature 359:162-165(1992).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=94140360; PubMed=8307570;			
RX	Plougaestel B., Zucman J., Peter M., Thomas G., Delattre O.;			
RA	"Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation.";			
RT	Genomics 18:609-615(1993).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	Zucman-Rossi J.;			
RC	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSTB=Placenta;			
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schlier G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stappleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ussid T.B., Tohyuki S., Carninci P., Prange C., Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A., Fehey J., Helton E., Kettlemen M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,			

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).  
 RN [5]  
 RX SEQUENCE OF 1-345 FROM N.A.  
 RX MEDLINE=97131501; PubMed=8975669; DOI=10.1006/geno.1996.0625;  
 RA Zucman-Rossi J., Legoux P., Thomas G.;  
 RT "Identification of new members of the Gas2 and Ras families in the  
 RT 22q12 chromosome region.";  
 RL Genomics 38:247-254(1996).  
 RN [6]  
 RP SEQUENCE OF 241-268 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95352541; PubMed=7542907;  
 RA Bhagavat T., Abe S., Nojima T., Yoshida M.C.;  
 RT "Molecular analysis of a t(11;22) translocation junction in a case of  
 RT Ewing's sarcoma.";  
 RL Genes Chromosomes Cancer 13:126-132(1995).  
 RN [7]  
 RP PHOSPHORYLATION SITE SER-266, AND PARTIAL SEQUENCE.  
 RX MEDLINE=98001723; PubMed=9341188; DOI=10.1074/jbc.272.43.27369;  
 RA Deloume J.C., Pritchard L., Delattre O., Storm D.R.;  
 RT "The proto-oncogene EWS binds calmodulin and is phosphorylated by  
 RT protein kinase C through an IQ domain.";  
 RL J. Biol. Chem. 272:27369-27377(1997).  
 RN [8]  
 RP SEQUENCE OF 128-158, 233-247, 268-324, 334-364, 393-439, 447-518 AND  
 RP 551-641, METHYLATION OF ARGININES, AND MASS SPECTROMETRY.  
 RX MEDLINE=1276345; PubMed=11278906; DOI=10.1074/jbc.M011446200;  
 RA Belyanskaya L.L., Gehrig P.M., Gehrig H.;  
 RT "Exposure on cell surface and extensive arginine methylation of Ewing  
 RT sarcoma (EWS) protein.";  
 RL J. Biol. Chem. 276:18661-18667(2001).  
 RN [9]  
 RP ALTERNATIVE SPLICING, AND RNA-BINDING.  
 RX MEDLINE=94366763; PubMed=8084618;  
 RA Ohno T., Ouchida M., Lee L., Gatalica Z., Rao V.N., Reddy B.S.P.;  
 RT "The EWS gene, involved in Ewing family of tumors, malignant melanoma  
 RT of soft parts and desmoplastic small round cell tumors, codes for an  
 RT RNA binding protein with novel regulatory domains.";  
 RL Oncogene 9:3087-3097(1994).  
 RN [10]  
 RP INTERACTION WITH SP1.  
 RX MEDLINE=9835009; PubMed=9660765; DOI=10.1074/jbc.273.29.18086;  
 RA Zhang D., Paley A.J., Childs G.;  
 RT "The transcriptional repressor ZFP1 interacts with and modulates the  
 RT ability of EWS to activate transcription.";  
 RL J. Biol. Chem. 273:18086-18091(1998).  
 RN [11]  
 RP CHARACTERIZATION.  
 RX MEDLINE=2039006; PubMed=10767297; DOI=10.1074/jbc.M002961200;  
 RA Li K.K.C., Lee K.A.W.;  
 RT "Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can  
 RT be cis-repressed by the EWS RNA-binding domain.";  
 RL J. Biol. Chem. 275:23053-23058(2000).  
 CC -1- FUNCTION: Might normally function as a repressor. EWS-fusion-  
 CC proteins (EFPs) may play a role in the tumorigenic process. They  
 CC may disturb gene expression by mimicking, or interfering with the  
 CC normal function of CTD-POLII within the transcription initiation  
 CC complex. They may also contribute to an aberrant activation of the  
 CC fusion protein target genes.  
 CC -1- SUBUNIT: Binds POLR2C, SP1, calmodulin and RNA. Interacts with  
 CC PRK2/PK2.  
 CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.  
 CC Relocates from cytoplasm to ribosomes upon PRK2/PK2 activation.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=EWS;  
 CC IsoId=001844-1; Sequence=Displayed;  
 CC Name=EWS-B;  
 CC IsoId=001844-2; Sequence=VSP\_005793;  
 CC Note=No experimental confirmation available;

CC TISSUE SPECIFICITY: Ubiquitous.  
 CC -1- DOMAIN: EWS activation domain (EAD) functions as a potent  
 CC activation domain in EFPs. EWS1 binds POLR2C but not POLR2E or  
 CC POLR2G, whereas the isolated EAD binds POLR2E and POLR2G but not  
 CC POLR2C. Cis-linked RNA-binding domain (RBD) can strongly and  
 CC specifically repress trans-activation by the EAD.  
 CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation  
 CC of Ser-266.  
 CC -1- DISBASE: Ewing's sarcoma is characterized by chromosomal  
 CC translocations t(11;22)(q24;q12) which involves EWS1 and FLI1,  
 CC t(7;22)(p22;q12) which involves EWS1 and ETV1, t(21;22)(q22;q12)  
 CC which involves EWS1 and ERG and t(9;22)(q22-31;q1-12) which  
 CC involves EWS1 and NR4A3.  
 CC -1- DISBASE: Involved in desmoplastic small round cell tumor (DSRCT)  
 CC through a chromosomal translocation t(11;22)(p13;q12) that  
 CC involves EWS1 and WT1.  
 CC -1- DISBASE: Malignant melanoma of soft parts (MMP), also known as  
 CC soft tissue clear cell sarcoma, is a rare tumor developing in  
 CC tendons and aponeuroses. It is associated with chromosomal  
 CC translocation t(12;22)(q13;q12) involving EWS1 and ATF-1.  
 CC -1- DISBASE: Involved in small round cell sarcoma through a  
 CC chromosomal translocation t(11;22)(p36.1;q12) that involves EWS1  
 CC and ZNF78.  
 CC -1- MISCELLANEOUS: EFPs arise due to chromosomal translocations in  
 CC which EWS1 is fused to a variety of cellular transcription  
 CC factors. EFPs are very potent transcriptional activators dependent  
 CC on the EAD and a C-terminal DNA-binding domain contributed by the  
 CC fusion partner. The spectrum of malignancies associated with EFPs  
 CC are thought to arise via EFP-induced transcriptional deregulation,  
 CC with the tumor phenotype specified by the EWS1 fusion partner and  
 CC cell type. Transcriptional repression of the transforming growth  
 CC factor beta type II receptor (TGF beta RII) is an important target  
 CC of the EWS-FLI1, EWS-ERG, or EWS-ETV1 oncogene.  
 CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the  
 CC absence, of calcium ion.  
 CC -1- SIMILARITY: Belongs to the RNP TET family.  
 CC -1- SIMILARITY: Contains 1 IQ domain.  
 CC -1- SIMILARITY: Contains 1 RanBP2-type zinc finger.  
 CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
 CC -1- DATABASE: NMR=Atlas Genet. CytoGenet. Oncol. Haematol.;  
 CC WWW=http://www.infobiogen.fr/services/chronocancer/Genes/EWS11085.html".  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EWS1: X66899; CAA47350.1; -  
 CC EWS1: X72990; CAA51488.1; -  
 CC EWS1: X72991; CAA51489.1; JOINED.  
 CC EWS1: X72992; CAA51489.1; JOINED.  
 CC EWS1: X72993; CAA51489.1; JOINED.  
 CC EWS1: X72994; CAA51489.1; JOINED.  
 CC EWS1: X72995; CAA51489.1; JOINED.  
 CC EWS1: X72996; CAA51489.1; JOINED.  
 CC EWS1: X72997; CAA51489.1; JOINED.  
 CC EWS1: X72998; CAA51489.1; JOINED.  
 CC EWS1: X72999; CAA51489.1; JOINED.  
 CC EWS1: X73000; CAA51489.1; JOINED.  
 CC EWS1: X73001; CAA51489.1; JOINED.  
 CC EWS1: X73002; CAA51489.1; JOINED.  
 CC EWS1: X73003; CAA51489.1; JOINED.  
 CC EWS1: X73004; CAA51489.1; JOINED.  
 CC EWS1: BC004817; AA004817.1; -  
 CC EWS1: Y08806; CAA70044.1; ALT\_INIT.  
 CC EWS1: AB016435; BAA31990.1; -  
 CC EWS1: Y07848; CAA69177.1; -  
 CC PIR; A49358; A49358.  
 CC HSSP; 095218; I002.  
 CC GeneW; HGNC:3508; EWS1.



DR H-InvDB: HIT0016349; --  
MIM: 133450; --

Query Match 100.0%; Score 3633; DB 1; Length 656;  
Best Local Similarity 100.0%; Pred. No. 2, 7e-148;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDSTYSQAQAAGGSAATTAQPTGTAQTTQAGGQSYGTGPTDVSYTAQTTAT 60  
DB 1 MASTDSTYSQAQAAGGSAATTAQPTGTAQTTQAGGQSYGTGPTDVSYTAQTTAT 60  
QY 61 YGQTAATATSGGPPPTGTTTPTAQAASQAPVQSGTGAVDTTTATVTTTQASVAASAYGT 120  
DB 61 YGQTAATATSGGPPPTGTTTPTAQAASQAPVQSGTGAVDTTTATVTTTQASVAASAYGT 120  
QY 121 QPAYPAVGGQPAATATPAPDGNKPTETSPQSSGTGYNPGLGYSQSNYSYQVPGSY 180  
DB 121 QPAYPAVGGQPAATATPAPDGNKPTETSPQSSGTGYNPGLGYSQSNYSYQVPGSY 180  
QY 181 MQPVTAPSPYPTSYSTQPTSTYDQSSYQONTYGQSSYGGQSSYGGQSSYGGQSSY 240  
DB 181 MQPVTAPSPYPTSYSTQPTSTYDQSSYQONTYGQSSYGGQSSYGGQSSYGGQSSY 240  
QY 241 PPGTGSYSQAPSPSYQSSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 300  
DB 241 PPGTGSYSQAPSPSYQSSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 300  
QY 301 GRRGGGFRDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
DB 301 GRRGGGFRDGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
QY 361 SAIVYQGLNDSVTLDDLADFFKCCGVYKAKRTGQPMTHIYLDKTKGPKGATVSYEDP 420  
DB 361 SAIVYQGLNDSVTLDDLADFFKCCGVYKAKRTGQPMTHIYLDKTKGPKGATVSYEDP 420  
QY 421 PTKAAVEMVDGDFQSSKTKVSLARKKPPMNSMRGGLPRBERGMPPLRGAGGPGGCP 480  
DB 421 PTKAAVEMVDGDFQSSKTKVSLARKKPPMNSMRGGLPRBERGMPPLRGAGGPGGCP 480  
QY 481 GGPWGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
DB 481 GGPWGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
QY 541 NCKKAPKPGGFLPPPPPPGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
DB 541 NCKKAPKPGGFLPPPPPPGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
QY 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
DB 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

## RESULT 2

Q96FE8 PRELIMINARY; PRT; 655 AA.  
ID Q96FE8  
AC Q96FE8; 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)  
DE Ewing sarcoma breakpoint region 1, isoform EMS (EMSR1 protein).  
GN Name=EMSR1;  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxId=9606;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph, and Skin;  
RX MEDLINE=2388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalek U., Smailus D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,  
RA Cole C.G., Goward M.B., Aguado B., Mallia M., Mokrab Y., Huckle E.J.,  
RA Beare D.M., Dunham I.;  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC011048; AAH11048.1; --  
DR EMBL: BC072442; AAH72442.1; --  
DR EMBL: CR456490; CAG30376.1; --  
DR HSSP: 095218; 1N02.  
DR GO: GO:0005634; C:nucleus; IEA.  
DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR InterPro: IPR011368; RNA-binding\_EWS.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR InterPro: IPR001876; Znf\_RangDP.  
DR Pfam: PF00641; ZF\_RanBP\_1.  
DR Pfam: PF00641; ZF\_RanBP\_1.  
DR SMART: SMART002101; RNA-binding\_EWS; 1.  
DR SMART: SMART00360; RRM; 1.  
DR SMART: SMART00547; Znf\_RBZ; 1.  
DR PROSITE: PS50102; RRM; 1.  
DR PROSITE: PS01358; ZF\_RanBP2\_1; 1.  
DR PROSITE: PS50159; ZF\_RanBP2\_2; 1.  
SQ SEQUENCE 655 AA; 68391 MW; B539ED1E98C601ED CRC64;

Query Match 99.6%; Score 3618.5; DB 2; Length 655;  
Best Local Similarity 99.8%; Pred. No. 1, 1e-147;  
Matches 655; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAAGGSAATTAQPTGTAQTTQAGGQSYGTGPTDVSYTAQTTAT 60  
DB 1 MASTDSTYSQAQAAGGSAATTAQPTGTAQTTQAGGQSYGTGPTDVSYTAQTTAT 60  
QY 61 YGQTAATATSGGPPPTGTTTPTAQAASQAPVQSGTGAVDTTTATVTTTQASVAASAYGT 120  
DB 61 YGQTAATATSGGPPPTGTTTPTAQAASQAPVQSGTGAVDTTTATVTTTQASVAASAYGT 120  
QY 121 QPAYPAVGGQPAATATPAPDGNKPTETSPQSSGTGYNPGLGYSQSNYSYQVPGSY 180  
DB 121 QPAYPAVGGQPAATATPAPDGNKPTETSPQSSGTGYNPGLGYSQSNYSYQVPGSY 180  
QY 181 MQPVTAPSPYPTSYSTQPTSTYDQSSYQONTYGQSSYGGQSSYGGQSSYGGQSSY 240  
DB 181 MQPVTAPSPYPTSYSTQPTSTYDQSSYQONTYGQSSYGGQSSYGGQSSYGGQSSY 240  
QY 241 PPGTGSYSQAPSPSYQSSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 300

Db	241	PPGTGYSQA	PSQ	YSQSS	YVQ	QSS	FRQ	DHSS	NGVY	GGSS	GGFS	GGSE	RSMS	GG	DNR	3.00																															
Qy	301	GRGRG	F	DRGMS	RGR	GG	RGRG	MS	AG	RGRG	F	NK	RGP	MD	B	GLD	GG	VP	VD	ED	SDN	3.66																									
Db	301	GRGRG	F	DRGMS	RGR	GG	RGRG	MS	AG	RGRG	F	NK	RGP	MD	B	GLD	GG	VP	VD	ED	SDN	3.59																									
Qy	361	SAIYV	GL	ND	SV	LD	LD	AD	F	FK	QCV	YV	K	N	K	T	Q	P	M	I	H	Y	LD	K	ET	G	K	DA	T	V	S	Y	ED	4.20													
Db	360	SAIYV	GL	ND	SV	LD	LD	AD	F	FK	QCV	YV	K	N	K	T	Q	P	M	I	H	Y	LD	K	ET	G	K	DA	T	V	S	Y	ED	4.19													
Qy	421	PTKAA	V	EN	F	D	K	D	F	O	G	S	K	L	K	T	V	S	L	A	R	K	P	P	N	S	M	F	G	L	P	P	B	G	R	M	P	P	L	R	G	F	G	F	G	P	4.80
Db	420	PTKAA	V	EN	F	D	K	D	F	O	G	S	K	L	K	T	V	S	L	A	R	K	P	P	N	S	M	F	G	L	P	P	B	G	R	M	P	P	L	R	G	F	G	P	4.79		
Qy	481	GGPMG	M	GG	GG	CD	R	GF	P	PR	G	R	G	S	R	G	N	S	G	G	A	N	O	H	A	G	M	O	C	N	P	C	C	G	N	O	N	A	M	A	M	T	E	C	5.40		
Db	480	GGPMG	M	GG	GG	CD	R	GF	P	PR	G	R	G	S	R	G	N	S	G	G	A	N	O	H	A	G	M	O	C	N	P	C	C	G	N	O	N	A	M	A	M	T	E	C	5.39		
Qy	541	NOC	K	A	P	E	G	L	P	P	P	P	P	G	D	R	G	R	G	P	G	M	R	G	S	G	L	M	D	R	G	F	P	G	M	F	R	G	R	G	CD	R	G	F	6.00		
Db	540	NOC	K	A	P	E	G	L	P	P	P	P	P	G	D	R	G	R	G	P	G	M	R	G	S	G	L	M	D	R	G	F	P	G	M	F	R	G	R	G	CD	R	G	F	5.99		
Qy	601	GGG	M	R	G	G	G	G	R	G	R	G	P	G	P	G	P	L	M	E	O	M	GG	R	G	R	G	R	G	P	R	M	D	G	E	H	R	O	E	R	R	D	P	Y	6.56		
Db	600	GGG	M	R	G	G	G	R	G	R	G	P	G	P	G	P	L	M	E	O	M	GG	R	G	R	G	R	G	P	R	M	D	G	E	H	R	O	E	R	R	D	P	Y	6.55			

RESULT 3	PRELIMINARY;	PRT;	661 AA.
096MX4			
AC	096MX4		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)		
DE	Hypothetical protein FLJ131747.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=14702039; DOI=10.1038/ng1285;		
RA	Ora T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,		
RA	Nakanatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Nakita H.,		
RA	Sekine M., Okeyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamanoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,		
RA	Murakami K., Uesada T., Iwayanagi T., Wagatsuma M., Shiratori A.,		
RA	Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikawa E., Omura Y.,		
RA	Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,		
RA	Ninomiya K., Iehishahi T., Yamashita H., Murakawa K., Fujimori K.,		
RA	Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,		
RA	Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta T., Kusano S.,		
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,		
RA	Togitsu S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,		
RA	Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,		
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shibahara N., Sano S.,		
RA	Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,		
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,		
RA	Hishigaki S., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Yamazaki M., Watanabe K., Kamegaki A., Itakura S., Fukusumi Y.,		
RA	Fujimori Y., Komiyama M., Taniguchi A., Tanigami A., Fujitara T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,		
RA	Kawabata A., Hikita T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Kaitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,		
RA	Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,		
RA	Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,		
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,		
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,		
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki S.,		
RT	"Complete sequencing and characterization of 21,243 full-length human		
RT	CDNA's."		

Query Match	Score	Length
Beat Local Similarity	98.9%	661
Matches	655	Indels 7, Gaps 2

QY	1	MASTDVSTYSQAAAQGGASATTAQPTQGGAAQTTAYAGQGSYGTGQPFVDVSTYTAQCTAT	60
Db	1	MASTDVSTYSQAAAQGGASATTAQPTQGGAAQTTAYAGQGSYGTGQPFVDVSTYTAQCTAT	60
QY	61	YGGTAATATSYGQPP-----TGTTTPTAPAYAGQPVQGGYGTGAVDTTATVTTTQASAYAA	114
Db	61	YGGTAATATATSYGQPPETVEGTSITGTTTPTAPAYAGQPVQGGYGTGAVDTTATVTTTQASAYAA	120
QY	115	QSATGTGPAYAYAGQBPAAATPAPTPROGKNKPTERSQPSQSSGTGYNQPSLGGYGQSSYSPQ	174
Db	121	QSATGTGPAPAYAGQBPAAATPAPTPROGKNKPTERSQPSQSSGTGYNQPSLGGYGQSSYSPQ	180
QY	175	VPGSYPMQPVTTAPPSYPTPTSYSSQTPTSYDQSSYSQNTYGPQSSYGOQSSSYGQ	234
Db	181	VPGSYPMQPVTTAPPSYPTPTSYSSQTPTSYDQSSYSQNTYGPQSSYGOQSSSYGQ	240
QY	235	QPPTSYPPQTCSSYQAPSYQSSSYGOQSSYFRQDHPSSMGVYGQSSGGSFSGPGENRSM	294
Db	241	QPPTSYPPQTCSSYQAPSYQSSSYGOQSSYFRQDHPSSMGVYGQSSGGSFSGPGENRSM	300
QY	295	SGPNNRGRGGRGFPDRCGMSRGRGGRGCGMSAABERGAFNKPQGGPMDGEPDLIDGPPVDP	354
Db	301	SGPNNRGRGGRGFPDRCGMSRGRGGRGCGMG - AABERGAFNKPQGGPMDGEPDLIDGPPVDP	359
QY	355	DESDNSAIIYQGLNDSTYLLDLADFPKQCGVYKMNKRTGQPMIHIYLDKKTGPKKGPAT	414
Db	360	DESDNSAIIYQGLNDSTYLLDLADFPKQCGVYKMNKRTGQPMIHIYLDKKTGPKKGPAT	419
QY	415	VSYEDPPTAAKAAVEFMFGKDFQFQSGKLKVSLARKKPPMNSMRGGLPPEBGRAMPPLRGGP	474
Db	420	VSYEDPPTAAKAAVEFMFGKDFQFQSGKLKVSLARKKPPMNSMRGGLPPEBGRAMPPLRGGP	479
QY	475	GGPGGPGCGPMGRMGGRGDRGFRPRGPRGSRGNPSCGGYVQHBADGWQCPNPGCGNPF	534
Db	480	GGPGGPGCGPMGRMGGRGDRGFRPRGPRGSRGNPSCGGYVQHBADGWQCPNPGCGNPF	539
QY	535	AMRETCNQCAKPKKEGLPPPPPPPGGDRGRCGGRGGRGGLMDGGPCGMPFRGGG	594
Db	540	AMRETCNQCAKPKKEGLPPPPPPPGGDRGRCGGRGGRGGLMDGGPCGMPFRGGG	599
QY	595	DRGGFRGGRGMDRGFGGGRGRRGCGPPLPMEQMGRRGGRGCGPKMDKGEHRQERRDR	654
Db	600	DRGGFRGGRGMDRGFGGGRGRRGCGPPLPMEQMGRRGGRGCGPKMDKGEHRQERRDR	659
QY	655	PY 656	
Db	660	PY 661	

## RESULT 4

ID	GENVA3	PRELIMINARY;	PRT;	656 AA.
AC	GENVA3;			
DT	05-JUL-2004	(TREMBLrel. 27, Created)		
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)		
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)		
DE	Ewing sarcoma breakpoint region 1.			
OS	Name=Ewari;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid11090;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RC	MEBLIN=22386857; PubMed=12477932; DOI=10.1073/pnae.242603899;			
RA	Strauberg R. L., Feringold E. A., Grouse L. H., Derge J. G.,			
RA	Klausner R. D., Collins F. S., Wagner L., Shemen C. M., Schler G. D.,			
RA	Altschul S. F., Zeeberg B., Buetow K. H., Schaefer C. F., Bhat N. K.,			
RA	Hopkins R. F., Jordan H., Moore T., Max S. I., Wang J., Hsieh F.,			
RA	Ditachenko L., Marustina K., Farmer A. A., Rubin G. M., Hong L.,			
RA	Stepleton M., Soares M. B., Bonaldi M. P., Casavant T. L., Scheetz T. E.,			
RA	Brownstein M. J., Ueda T. B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S. S., Loquellano N. A., Peters G. J., Abramson R. D., Mollay S. J.,			
RA	Bosak S. A., McEwen P. J., McKernan K. J., Malek J. A., Gunaratne P. H.,			
RA	Richards S., Worley K. C., Hale S., Garcia A. M., Gay L. J., Hulyk S. W.,			
RA	Villation D. K., Muzny D. M., Sodergren E. J., Lu X., Gibbs R. A.,			
RA	Fahney J., Helton E., Kettelman M., Madan A. C., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A. C., Shevchenko Y., Bouffard G. G.,			
RA	Blakeley R. W., Touchman J. W., Green E. D., Dickson M. C.,			
RA	Rodriguez A. C., Grimwood J., Schmutz J., Myers R. M., Butterfield Y. S.,			
RA	Krzywniak M. I., Skalska U., Smallus D. E., Scherch A., Schein J. E.,			
RA	Jones S. U., Mair M. A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[2]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RC	Director MGC Project;			
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.			
DR	GEO: BC068226; AAB68226.1; -			
DR	GO: GO:0005634; C:nucleus; IDA.			
DR	InterPro: IPR011368; RNA-binding_EWS.			
DR	InterPro: IPR000504; RNA_rec_mot.			
DR	InterPro: IPR001876; znf_RangDP.			
DR	Pfam: PF00076; RRM_1; 1.			
DR	Pfam: PF00641; zF-RanBP; 1.			
DR	PIRSF: PIRSF02101; RNA-binding_EWS; 1.			
DR	SMART: SM00360; RRM; 1.			
DR	SMART: SM00547; znf_RBZ; 1.			
DR	PROSITE: PS0102; RRM; 1.			
DR	PROSITE: PS01358; zF_RanBP2_1; 1.			
DR	PROSITE: PS0199; zF_RanBP2_2; 1.			
SO	SEQUENCE 656 AA; 68549 MW; 5DP018A22C8C1D0B CRC64;			
Query Match	98.5%; Score 3579; DB 2; Length 656;			
Best Local Similarity	97.9%; Pred. No. 5,6e-146;			
Matches 642; Conservative	9; Mismatches 5; Indels 0; Gaps 0			
CY	1 MASTDYSTSYGSAAGGYSAYTAQPTGQYAGTTQYAGGQSYGTYGQPTDVSYTQAGTTAT			60
DB	1 MASTDYSTSYGSAAGGYSAYTAQPTGQYAGTTQYAGGQSYGTYGQPTDVSYTQAGTTAT			60
CY	61 YGQRAVAATSYGQPPFGYSTPTPAQAQYSPGVQGYGTYDSTATVATTTQASAAAGTAATGT			120
DB	61 YGQRAVAATSYGQPPFGYSTPTPAQAQYSPGVQGYGTYDSTATVATTTQASAAAGTAATGT			120
CY	121 QPAPVAYGQCPAAATAPTRPQDGNKRTETSQPOSSTGCVNQPLAGQGSNYSYPOVPGSYF			180
DB	121 QPAPVAYGQCPAAATAPTRPQDGNKRTETSQPOSSTGCVNQPLAGQGSNYSYPOVPGSYF			180
CY	181 MQPVTAPSPYPTTSYSTQPTSYDQSSYSQONTYQPPSSYGQSSYGGQSPPTSY.240			

Db	181	MQPTTAPPSYPTSYSSSSQPRSYVSSQNTYTGSSSYGQSSYQGQSSYGGQPTSY	240
Qy	241	PPQTGSYSQABSYQSQQSSSYGQSSYFRQDHPSSMGVYQGESGFGPGENRSMGPDNR	300
Db	241	PPQTGSYSQABSYQSQQSSSYGQSSYFRQDHPSSMGVYQGESGFGPGENRSLSGPDNR	300
Qy	301	GRGGRGPDRCGMSNGRGGGRGGMGMSAGRGGRFPKPGPMDRGGDLDLGPVVDDESDN	360
Db	301	GRGGRGPDRCGMSNGRGGGRGGMGMSAGRGGRFPKPGPMDRGGDLDLGPVVDDESDN	360
Qy	361	SAIYVQGLNDVTLDDLADFPKQCGVYMNKRQTQCPMIIHYLDKETGPKKDAVSYEDP	420
Db	361	SAIYVQGLNDVTLDDLADFPKQCGVYMNKRQTQCPMIIHYLDKETGPKKDAVSYEDP	420
Qy	421	PTAKAAVEWFPDGKDFQGSKLKVLARKKPPMNSMRGLPERRGMPPLRGGFGPGR	480
Db	421	PTAKAAVEWFPDGKDFQGSKLKVLARKKPPMNSMRGMPERRGMPPLRGGFGPGR	480
Qy	481	GGPMGRMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR	540
Db	481	GGPMGRMGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR	540
Qy	541	NOCKAPPEGRFLPPPPPPGGDRGRGPGCGRRGGRGGLMDRGCGMGFRGGRGGDRGFR	600
Db	541	NOCKAPPEGRFLPPPPPPGGDRGRGPGCGRRGGRGGLMDRGCGMGFRGGRGGDRGFR	600
Qy	601	GGRGMDRGFGFGGGRGGRGPGCPGLMEOMGRGGRGGRGPGCMKGEHROERDRPY	656
Db	601	GGRGMDRGFGFGGGRGGRGPGCPGLMEOMGRGGRGGRGPGCMKGEHROERDRPY	656

RESULT 5

EWS\_MOUSE

ID

EWS\_MOUSE

STANDARD;

PRT;

655 AA.

AC

061545;

DT

01-NOV-1997 (Rel. 35, Created)

DT

01-NOV-1997 (Rel. 35, Last sequence update)

DT

25-OCT-2004 (Rel. 45, Last annotation update)

DE

RNA-binding protein EWS.

GN

Name=Ewst1; Synonym=EWS, Ewsh;

OS

Mus musculus (Mouse).

OC

Eukaryota; Euteleostei; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX

NCBI\_TaxID=10090;

RM

[1]

RP

SEQUENCE FROM N.A.

RC

TISSUE=Testis;

RX

MEBLINE=95130099; PubMed=7829090;

RA

Plougastel B., Matei M.-G., Thomas G., Delattre O.;

RT

"Cloning and chromosome localization of the mouse Ews gene.";

RL

Genomics 23:278-281(1994).

RL

-1- FUNCTION: Might function as a repressor (By similarity).

CC

-1- SUBUNIT: Binds RNA, POLR2C, SFI and calmodulin. Interacts with

CC

PTX2B (By similarity).

CC

-1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.

CC

-1- Relocates from cytoplasm to ribosomes upon PTX2B/FAK2 activation

CC

(By similarity).

CC

-1- PM: Phosphorylated; calmodulin-binding inhibits phosphorylation

CC

of Ser-266 (By similarity).

CC

-1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the

CC

absence, of calcium ion (By similarity).

CC

-1- SIMILARITY: Belongs to the RNP TET family.

CC

-1- SIMILARITY: Contains 1 IQ domain.

CC

-1- SIMILARITY: Contains 1 RanzB-type zinc finger.

CC

-1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.

CC

-----

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC

the European Bioinformatics Institute. There are no restrictions on its

CC

use by non-profit institutions as long as its content is in no way

CC

modified and this statement is not removed. Usage by and for commercial

CC

entities requires a license agreement (See <http://www.isb.ch/announce/>

or send an email to license@sb-stb.ch).

CC -----

DR EMBL: X79233; CA55815.1; -

DR PIR: A55726; A55726.

DR HSSP: O95218; IM02.

DR MGD; MGI:99960; Ewsh.

DR GO: GO:0005634; C:nucleus; IDA.

DR InterPro: IPR000048; IO\_region.

DR InterPro: IPR000504; RNA\_rec\_mot.

DR InterPro: IPR001876; Znf\_RandP.

DR Pfam: PF00076; RRM\_1; 1.

DR Pfam; PF00641; zf-RandP; 1.

DR PIRSF: PIRSF002101; RNA-binding\_EWS; 1.

DR PROSITE; PSS0096; IO; FALSE\_NEG.

DR PROSITE; PSS0102; RRM; 1.

DR PROSITE; PSS0138; ZF\_RANBP2\_1; 1.

DR PROSITE; PSS0199; ZF\_RANBP2\_2; 1.

DR Calmodulin-binding; Metal-binding; Methylation; Nuclear protein;

KW Phosphorylation; Repeat; Repressor; RNA-binding;

KM Transcription regulation; zinc; zinc-finger.

FT DOMAIN 1 285 EAD (Gln/Pro/Thr-rich).

FT DOMAIN 256 285 IO.

FT DOMAIN 360 446 RNA-binding (RRM).

FT ZN\_FING 517 548 RANBP2-type.

FT 8 285 31 X approximate tandem repeats.

FT REPEAT 17 27 2.

FT REPEAT 28 34 3.

FT REPEAT 35 42 4.

FT REPEAT 43 50 4.

FT REPEAT 51 59 6.

FT REPEAT 60 68 7.

FT REPEAT 69 75 8.

FT REPEAT 76 84 9.

FT REPEAT 85 91 10.

FT REPEAT 92 110 11.

FT REPEAT 111 116 12.

FT REPEAT 117 125 13.

FT REPEAT 126 156 14.

FT REPEAT 157 163 15.

FT REPEAT 164 170 16.

FT REPEAT 171 177 17.

FT REPEAT 178 188 18.

FT REPEAT 189 193 19.

FT REPEAT 194 201 20.

FT REPEAT 202 206 21.

FT REPEAT 207 212 22.

FT REPEAT 213 218 23.

FT REPEAT 219 224 24.

FT REPEAT 225 230 25.

FT REPEAT 231 238 26.

FT REPEAT 239 245 27.

FT REPEAT 246 252 28.

FT REPEAT 253 259 29.

FT REPEAT 260 276 30.

FT REPEAT 277 285 31.

FT DOMAIN 300 339 Arg/Gly/Pro-rich.

FT DOMAIN 453 512 Arg/Gly/Pro-rich.

FT DOMAIN 558 639 Arg/Gly/Pro-rich.

FT MOD\_RES 300 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 302 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 304 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 309 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 314 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 317 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 321 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 454 454 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 463 463 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 470 470 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 489 489 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 493 493 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 499 499 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 502 502 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 505 505 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 562 562 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 564 564 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 571 571 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 574 574 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 580 580 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 588 588 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 591 591 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 595 595 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 599 599 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 602 602 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 606 606 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 614 614 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 632 632 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 635 635 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 655 AA; 68418 MW; 50735EDB54247D69 CRC64; Phosphoserine (By PKC) (By similarity).

SO SEQUENCE

Query Match 98.3%; Score 3571.5; DB 1; Length 655;

Best Local Similarity 98.0%; Pred. No. 1.2e-145;

Matches 643; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAAGGSAVTAOPTQGAOTTAAGGSSYGVGQPTDVSYQAQTAT 60

DB 1 MASTDSTYSQAQAAGGSAVTAOPTQGAOTTAAGGSSYGVGQPTDVSYQAQTAT 60

QY 61 YGQTAATVATSGOPTGTTPTAQAAYSQVQVGTGAVDTTATVTTTQAASVAAGAT 120

DB 61 YGQTAATVATSGOPTGTTPTAQAAYSQVQVGTGAVDTTATVTTTQAASVAAGAT 120

QY 121 QPAYPAVGGQPAATPATRQDGNKPTETQPOSSTGCTGNQPSIGYQGSNTSYQVPGSTP 180

DB 121 QPAYPAVGGQPAATPATRQDGNKPTETQPOSSTGCTGNQPSIGYQGSNTSYQVPGSTP 180

QY 181 MOPVTPAPSPYPTPSYSTQPTSDSSVQONTYGPSSYGGQSSYGGQSSYGGQSPPTSY 240

DB 181 MOPVTPAPSPYPTPSYSTQPTSDSSVQONTYGPSSYGGQSSYGGQSSYGGQSPPTSY 240

QY 241 PPTGSGSQAAPSOYSGQSSSYGQSSFRQDHSMSGVYQESGGFSGPENRSMGSPDNR 300

DB 241 PPTGSGSQAAPSOYSGQSSSYGQSSFRQDHSMSGVYQESGGFSGPENRSMGSPDNR 300

QY 301 GGRGGFDRGMSRGRGGGGRGMSAGRGGPNKFGGPMDEGPDLDLGPVDPDEDSDN 360



Db	360	SAIYVGGSLNDNTWLDLDLADAFKQCGVVMKMKRTGQPMHIIYLDKETGPKGADATVSIEDP	418
Qy	421	PTKAAVEMPRDGDPCGSKLYSLARKKPEPMNSMNGSLPREEGKGMPPRLRGSGGSGGR	480
Db	420	PTAKAAVEMPRDGDPCGSKLYSLARKKPEPMNSMNGSLPREEGKGMPPRLRGSGGSGGR	479
Qy	481	GGPMGMRGSGDRGSGPRPGRGSGRGNPSGGGNYQHPAGMGQCPNPBGCGNPFAMFTEC	540
Db	480	GGPMGMRGSGDRGSGPRPGRGSGRGNPSGGGNYQHPAGMGQCPNPBGCGNPFAMFTEC	539
Qy	541	NCKAKPEBSGFLPPPPPPFGSDRGKSGPCGMNGHNGSLMDNGFRQGMFRGSGGDRGSGFR	600
Db	540	NCKAKPEBSGFLPPPPPPFGSDRGKSGPCGMNGHNGSLMDNGFRQGMFRGSGGDRGSGFR	599
Qy	601	GGGMRGSGGCGRRGSGPCGPPRLMBEOMGSGRRGSGPGMNDGGEHROEERDDPY	656
Db	600	GGGMRGSGGCGRRGSGPCGPPRLMBEOMGSGRRGSGPGMNDGGEHROEERDDPY	655

## RESULT 7

ID	Q66MN4	PRELIMINARY;	PRT;	600 AA.			
AC	Q66MN4						
DT	01-DEC-2001	(TrEMBLrel. 19, Created)					
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)					
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)					
DE	Hypotheetical protein FLJ32119.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	PubMed=14702039; DOI=10.1038/ng1285;						
RA	Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,						
RA	Makametsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,						
RA	Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,						
RA	Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,						
RA	Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,						
RA	Sudo H., Hosobuchi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,						
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,						
RA	Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,						
RA	Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,						
RA	Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,						
RA	Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hottu T., Kusano J.,						
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,						
RA	Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,						
RA	Moshalino K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,						
RA	Yoshikawa Y., Matsumura H., Ichihara T., Shiohara N., Sano S.,						
RA	Moriya S., Momiyama H., Saich N., Takami S., Teraishi Y., Suzuki O.,						
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,						
RA	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,						
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,						
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,						
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,						
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,						
RA	Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,						
RA	Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,						
RA	Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,						
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,						
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamashita R.,						
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;						
RT	"Complete sequencing and characterization of 21,243 full-length human cDNAs."						
RL	Nat. Genet. 36:40-45(2004).						
DR	EMBL; AK056681; BAB71252.1; .						
DR	HSSP; O95218; INO2.						
DR	GO; GO:0005634; C:nucleus; IEA.						
DR	GO; GO:0003676; P:nucleic acid binding; IEA.						
DR	GO; GO:0008270; P:zinc ion binding; IEA.						
DR	InterPro; IPR011368; RNA-binding_EMS.						
DR	InterPro; IPR000504; RNA_rec_mec.						
DR	InterPro; IPR001876; Znf_RangDP.						

DR Pfam; PF00076; RRM\_1; 1.  
DR Pfam; PPO641; zf\_RanBP; 1.  
DR PIRSF; PIRSF02101; RNA-binding\_EWS; 1.  
DR SMART; SM00360; RRM; 1.  
DR SMART; SM00547; znp\_RBZ; 1.  
DR PROSITE; PS0103; RRM; 1.  
DR PROSITE; PS01358; ZF\_RANBP2\_1; 1.  
DR PROSITE; PS01199; ZF\_RANBP2\_2; 1.  
DQ SEQUENCE 600 AA; 62478 MW; 1572CEBF6C023912 CRC64;

Query Match	90.24;	Score 3277;	DB 2;	Length 600;
Best Local Similarity	91.34;	Pred. No. 4.4e-13;		
Matches 599;	Conservative	0;	Mismatches 1;	Indels 56;
				Gaps 12.

Qy	1	MASTDYSTYSOAAAOQSSVYTAQPTGQVACTTAYAGQSSYGTGVPDVSYTAQACTAT	60
Db	1	MASTDYSTYSOAAAOQSSVYTAQPTGQVACTTAYAGQSSYGTGVPDVSYTAQACTAT	60
Qy	61	YQGTAYVTSYGQPPTYGTTPTTAPQAYSQPVQGYGTGAYDITTAATVTTTQASVAAQASVGT	120
Dy	61	YQGTAYVTSYGQPPTYGTTPTTAPQAYSQPVQGYGTGAYDITTAATVTTTQASVAAQASVGT	120
Qy	121	QPAYPAVGOQPAATAPTRPDQGNKPTESQPSQSGTGYNQSLVYGGSNVSYPOVPGSY	180
Dy	121	QPAYPAVGOQPAATAPTRPDQGNKPTESQPSQSGTGYNQSLVYGGSNVSYPOVPGSY	180
Qy	181	MQPYTAPPSYPTSTYSSTOPTSYDQSSYQONTYGQPSYVGQSSYGOQSSYGOQPPTSY	240
Dy	136	-----PTSYSTOPTSYDQSSYQONTYGQPSYVGQSSYGOQSSYGOQPPTSY	184
Qy	241	PPQYSGTSOAPSOYSQOQSSSYGQSSSTRQDHPSSMSYTYGGSGGFSQPSGEMRSMSPBNR	300
Dy	185	PPQYSGTSOAPSOYSQOQSSSYGQSSSTRQDHPSSMSYTYGGSGGFSQPSGEMRSMSPBNR	244
Qy	301	GRGRGGDRGMSYGRGGRGGMKSAGBERGFFMKPGAPMDEGDDLQGPVYDDEBBDN	360
Dy	245	GRGRGGDRGMSYGRGGRGGMKSAGBERGFFMKPGAPMDEGDDLQGPVYDDEBBDN	304
Qy	361	SAIYVQGLNDSVTLDDLADFPKQGGVYKMKRTGQPMIHYLDKXETGPKKDATVSYEDP	420
Dy	305	SAIYVQGLNDSVTLDDLADFPKQGGVYKMKRTGQPMIHYLDKXETGPKKDATVSYEDP	364
Qy	421	PTAAABAEWDDGXDFQSSKLVSLAKKPPNNMSRGLPPEEGGMPPPLVGGPQGP	480
Dy	365	PTAAABAEWDDGXDFQSSKLVSLAKKPPNNMSRGLPPEEGGMPPPLVGGPQGP	424
Qy	481	GGPMRWGGRGSDGCGPPPRGPRSGRANPSGGGVQRAAGMOCPNPGCQNONPAMRTYC	540
Dy	425	GGPMRWGGRGSDGCGFPFRGPRSGRANPSGGGVQRAAGMOCPNPGCQNONPAMRTYC	484
Qy	541	NOCKAPRDEGFLPPPPPPGADRGGRGGGMRGGRGGLMDRGFGVGFMRGGRGDDGGR	600
Dy	485	NOCKAPRDEGFLPPPPPPGADRGGRGGGMRGGRGGLMDRGFGVGFMRGGRGDDGGR	544
Qy	601	GGRMDDGGCGGRGGRGPGAPPGGLMOMGGRGGRGPGCMDDGGEHRQERRDPY	656
Dy	545	GGRMDDGGCGGRGGRGPGAPPGGLMOMGGRGGRGPGCMDDGGEHRQERRDRY	600

## RESULT 8

ID	PRELIMINARY	PRT	674 AA.
OE33NO			
AC	QEP3NO		
DT	05-JUL-2004 (TREMBLrel. 27, Created)		
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)		
DE	Hypotheetical protein MGC76258.		
GN	Name=MGC76258		
OS	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxId=8364;		

360 SKPGGLDDGGDLDLGPMPPLPLPLDLDELEDESSTIYYQGLNDVYAEIYDFPKHCGDV 419

388 KMNRRGTGQPMHVIYLDKETGKPKKDAATVSYEDPPTAKAAVEMFGKDFQSGSLTKVSLARK 447

420 KINRTGTEPLVNLMDMETGPKKDDGVSPEDPSATATIELCGKDLNGKAVKVLARK 479

448 KPMNMSMRGGLPPREGGMPPLPLRGPGGPGGPGMGMWGRGGRGDRGFPDRPGSRG 507

480 KSLILGSMRGGSLLDNDNGQPEPLR-----GGPMGRIGGRGGRGGRGFMFRGPRGPRG 530

508 NPSSGGVVOHRAHGMOCNPPCCGNQNTAMRTKCNQCCAPKREGLP-PPPPPPGDRGGRG 566

531 SPV-SGNVQHAAGMOCNPPCCGNQNTAMRTKCNQCCAPKRGD--PPPPPPGDEGRERG 587

567 GPGGMRGGRGGGLMRGGGPGGFRGGRGDRGDRGFRGGRGMDRGGFGGGRRGGPGGPPGGLM 626

588 GP-GMRGR-GLMNRGVP-GMFRGGRGDRGGRFP-GGMMDR-GFGGGRRGPPPPGGLM 642

627 EQMGRRRG--RGSPGKMDXGHEORRDRPY 656

643 EQLGRIQGRGGRGPGSPKMDXSEHQRERERY 674

RESULT 9

06DJ04 PRELIMINARY; PRT; 673 AA.

06DJ04

AC 06DJ04; 25-OCT-2004 (TREMBlrel. 28, Created)

DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)

DE Hypothetical protein MGCF76258.

GN Name=MGCF76258;

OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

CC Amphibia; Batrachia; Anura; Mesodactrychia; Pipidea; Pipidae;

CC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8364;

OX [1]

RN RN

RP SEQUENCE FROM N.A.

RP TISSUE=Embryo;

RA MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RX Strausberg R.L., Fingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,

RA Stepien L.M., Soares W.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udén T.B., Toshitsuyuki S., Caminici P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Boeak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulik S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzyanek M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,

RA Jones S.U., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RL and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP [2]

RP SEQUENCE FROM N.A.

RP TISSUE=Embryo;

RA Klein S., Gerhard D.S.;

RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.

RL EMBL: BC075120; AAH75120.1; -

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003676; F:nucleic acid binding; IEA.

DR GO: GO:0008270; F:zinc ion binding; IEA.

DR InterPro: IPR011368; RNA-binding EMS.

DR InterPro: IPR011368; RNA-binding EMS.





QY 228 QGSSY-----GQPTSYSPONGSYSQ-APSOYSGQSSS-----YGOQ-----SSFR 268  
 Db 197 QGCGYQOTPPQQQAPSSSTAPSSGSGYGSASQYGGQSTGGGYSGYKPPNQGSYR 256  
 QY 269 QDHPSSMGVYQBSGFGSPGENRSMSPGNRGRGPGFPGGMSRGG--RGG-GRGGMG 325  
 Db 257 PDHQNGGGVYGGPSGGYGGPGDEGRGMCGENRGRGGRGPGFRGMWRGCGGMRGGSRSRGMG 316  
 QY 326 SAERGGFNNKPGGPMDBGPDLIDGPVPDPEDSDNSAIYVQGLNDSYTLDDLADFFKQCG 385  
 Db 317 IADRRGGFSPKPG-----DGEAGAPBEO--DSENSTIYTGLENTALEVDFPFHSG 368  
 QY 386 VVGNKRTGQPMHIYLDKETGPKPDATYSEDPPEPAKAAMVFDKDFQSGSLKXSLA 445  
 Db 369 IIRINKRTGPAVNIYTDKDTGPKDATTSTYEPPEAKAAVEMFDKDFQSKKLKXSLA 428  
 QY 446 RKPPNMSNRGGLPPRGRGMPPLLRGPGGPGGPGGPMRGMG-GRGDRGPGPR-GR 503  
 Db 429 RRRPMNMGMRGM-----PMRGDRGMGMGRGMGRGMGRGMGRGMGRGMGRGMGR 478  
 QY 504 G-SRGNPSGGGANTVQHRAGDMQCPNPGCGGNONFAMRTCCNOCKAPKEPFLPPPPPPG 562  
 Db 479 GMRGGGPT-CCNNQORAGDMQCPNAGCGGNONFAMRMCNOCCKAPKEPFGPPPF-PPG 536  
 QY 563 RGRGPGGMRGGRGLMDRGGPGGM--FRGGRGDRGGRGGRGMRGGRGGRGGRGGRG 620  
 Db 537 RGGGPGGMRGGRG--MDRGGPGGPGGFRGGRGDRGGRGFR-GRGMDGFGG--RGRGG 590  
 QY 621 PPGPLMEOMGRRGRGGRGPKMD-KGEHROERRDRPY 656  
 Db 591 PF---MDMGRRGRGMGPBGKMDKDRDRRERY 624

RESULT 11  
 Q9CRSS PRELIMINARY; PRT; 333 AA.  
 AC Q9CRSS;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
 DE Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched  
 DE library, clone:330002D11 product:Ewing sarcoma homolog, full insert  
 DE sequence. (fragment).  
 GN Name=Ewert1, Synonym=Ewbh;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
 NC NCBT\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.,  
 RT "High-efficiency full-length cDNA cloning.",  
 RL Mech. Enzymol. 303:19-44(1999).  
 RU Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;

RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.",  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Katsunari T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsuno H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa K., Ozawa K., Tanaka T., Matsumura S., Kawai J.,  
 RA Yoneda Y., Ishikawa T., Tanaka K., Tanaka T., Matsumura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.",  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RA Aachai J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arai K., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Iwatsuki K., Ishii Y., Itoh M., Izawa M., Kasakawa T., Kato H.,  
 RA Kawai J., Kojima Y., Kono H., Koda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka M.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK014366; BAB29301.1; -.  
 DR HSSP: O95218; 1N02.  
 DR MCD; MG1:99960; Ewert1.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR001876; Znf\_RanGDP.  
 DR Pfam; PF00076; RRM\_1; 1.  
 DR Pfam; PF00641; Zf\_RanBP; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR SMART; SM00547; Znf\_RBZ; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS01358; ZF\_RANBP2\_1; 1.  
 DR PROSITE; PS50199; ZF\_RANBP2\_2; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 333 AA; 34617 MW; CAEDBCA353FF2P07 CRC64;  
 Query Match 51.6%; Score 1874; DB 2; Length 333;  
 Best Local Similarity 98.5%; Pred. No. 3.1e-73;  
 Matches 328; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 324 MSGAGRGGPNKPGGPMDBGPDLIDGPVPDPEDSDNSAIYVQGLNDSYTLDDLADFFKQ 383  
 Db 1 LGSAGRGGPNKPGGPMDBGPDLIDGPVPDPEDSDNSAIYVQGLNDSYTLDDLADFFKQ 60  
 QY 384 CGVYNNKRTGQPMHIYLDKETGPKPDATYSEDPPEPAKAAMVFDKDFQSGSLKXYS 443  
 Db 61 CGVYNNKRTGQPMHIYLDKETGPKPDATYSEDPPEPAKAAMVFDKDFQSGSLKXYS 120  
 QY 444 LARKPPNMSNRGGLPPRGRGMPPLLRGPGGPGGPGGPMRGMGRGGRGGRGGRGGR 503  
 Db 121 LARKPPNMSNRGGMPPRGRGMPPLLRGPGGPGGPGGPMRGMGRGGRGGRGGRGGR 180  
 QY 504 GSRGNPSGGGANTVQHRAGDMQCPNPGCGGNONFAMRTCCNOCKAPKEPFLPPPPPPG 563  
 Db 181 GSRGNPSGGGANTVQHRAGDMQCPNPGCGGNONFAMRTCCNOCKAPKEPFLPPPPPPG 240  
 QY 564 GRGPGGMRGGRGLMDRGGPGGM--FRGGRGDRGGRGGRGMRGGRGGRGGRGGRGPG 623



RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzyzanski M.I., Skalka U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC000527; AAH00527.1; -  
 SQ SEQUENCE 354 AA; 37620 MW; AE4B8FCDF458390B CRC64;

Query Match 48.3%; Score 1753; DB 2; Length 354;  
 Best Local Similarity 99.1%; Pred. No. 5e-68; Mismatches 3; Indels 0; Gaps 0;  
 Matches 346; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MASTDSTYSQAQAQOQSAATTAQPTQGYAQTTOAYGQOQSYGTGPTDVSYTOAQTAT 60  
 DB 1 MASTDSTYSQAQAQOQSAATTAQPTQGYAQTTOAYGQOQSYGTGPTDVSYTOAQTAT 60  
 QY 61 YGCTAVATSYGPPPTGTTTTPAPQAVSQPVQYGTGAYDTTAAVTTTQAASVAAQSYGT 120  
 DB 61 YGCTAVATSYGPPPTGTTTTPAPQAVSQPVQYGTGAYDTTAAVTTTQAASVAAQSYGT 120  
 QY 121 QPAPYVYGGOPATATPRPDGKPKPTETSPQSTGYNPQSYGQSNYSYQVQVGSYP 180  
 DB 121 QPAPYVYGGOPATATPRPDGKPKPTETSPQSTGYNPQSYGQSNYSYQVQVGSYP 180  
 QY 181 MOPVTAPSPYPTSYSTOPTSYDQSSYSGQNTYGPSSYGGQSSYGGQSPPTSY 240  
 DB 181 MOPVTAPSPYPTSYSTOPTSYDQSSYSGQNTYGPSSYGGQSSYGGQSPPTSY 240  
 QY 241 PPTGTSYGAQPSQYSGQSSYGGQSSFRQDHPSSMGVYGGESGFGSGENRSMGPDNR 300  
 DB 241 PPTGTSYGAQPSQYSGQSSYGGQSSFRQDHPSSMGVYGGESGFGSGENRSMGPDNR 300  
 QY 301 GGRGGFDRGMSRGRGGRGGGMSAGB 329  
 DB 301 GGRGGFDRGMSRGRGGRGGGMSAGB 329

RESULT 14  
 0803E3 PRELIMINARY; PRT; 578 AA.

AC 0803E3;  
 DT 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Zgc:55864 protein.  
 GN Name: fuel; Synonym: zgc:55864; (Danio rerio).  
 OS Brachydanio rerio (Zebrafish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_Taxid:7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AB; TISSUE=whole body;  
 RX MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Klausner R.D., Collins F.S., Wagner K.H., Shemen C.M., Schlier G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Dlatshenko L., Marutina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Urdin T.B., Tothlyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,...

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzyzanski M.I., Skalka U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AB; TISSUE=whole body;  
 RA Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC044518; AAH44518.1; -  
 DR ZFIN; ZDB-GENE-030131-1600; fuel.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR001876; ZnF\_RanBP.  
 DR Pfam; PF00076; RRM\_1; 1.  
 DR Pfam; PF00641; zf\_RanBP; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR SMART; SM00547; ZnF\_RBZ; 1.  
 DR PROSITE; PS0102; RRM; 1.  
 DR PROSITE; PS01358; ZF\_RANBP2\_1; 1.  
 DR PROSITE; PS0199; ZF\_RANBP2\_2; 1.  
 SQ SEQUENCE 578 AA; 60625 MW; E959AACAD7004CD8 CRC64;

Query Match 43.5%; Score 1579; DB 2; Length 578;  
 Best Local Similarity 50.1%; Pred. No. 2.1e-60; Mismatches 141; Indels 136; Gaps 26;  
 Matches 343; Conservative 65; Mismatches 141; Indels 136; Gaps 26;

QY 1 MAS-TDYSTYSQAQAQOQSAATTAQPTQGYAQTTOAYGQOQSYGTGPTDVSYTOAQTAT 56  
 DB 1 MASVTYSSYNAQAQOQSYGTYAQTQGYAQTTOAYGQOQSYGTGPTDVSYTOAQTAT 60  
 QY 57 TTA-TYQTAATVATSYGQ-----PPTGTTTTPAPQAVSQPVQYGTGAYDTTAAVTTT 108  
 DB 61 PSAGAYAAQOQGYSTYGAQAATAAATAAATAAATGTP-QPQAYTQPAQSYGASVYGTSTA-APAA 118  
 QY 109 QASVAAQSAATTAQPTQGYAQTTOAYGQOQSYGTGPTDVSYTOAQTAT 168  
 DB 119 QASVGSQPGYSTQPAVSYGQOPASA----- 145  
 QY 169 NYSYQVQVGSYPMQPTAPPSPYPTSYSTOPTSYDQSSYSGQNTYGPSSYGGQSSYGGQSPPTSY 227  
 DB 146 -----PQSYASASQPAVYQSAVQPAQSYGQOQSYGTGPTDVSYTOAQTAT 182  
 QY 228 -QOQSSYG-----QOQPTSYPTQTSYGAQPSQYSGQSSYGGQSSFRQDHPSSMGV 277  
 DB 183 QOQSAVGGQOQPPQHOQOQGPAAVPPQ-GSSSYAQTYGQOQSA-----PQNDYQOQPNYS--- 235  
 QY 278 YGQSGSGFGSG--RGRMSRSGPDRNGRGRGPRGMSRGRGGRGGGMSAGBGRGPFNK 335  
 DB 236 YSG--GGVSGYVPGSGRG---GYQGGRGDYGDRGPR--GRMGWGKGWGIADRGGPFNK 287  
 QY 336 PGQPMQEGPDLDPGVVDDEBDSNGAIYVQGLNDSVTLDDLADFPKQGVVGMKRTGQ 395  
 DB 288 PGQPMQEGPDLDPGVVDDEBDSNGAIYVQGLNDSVTLDDLADFPKQGVVGMKRTGQ 345  
 QY 396 PMIHLYDKETGPKKDATVSYEDPTAAVAWEFPGDKFOGSKLKVSLARKKPPNMSNR 455  
 DB 346 PAINTYTDKSGPKKDATVSYEDPTAAVAWEFPGDKFOGSKLKVSLARKKPPNMSNR 405  
 QY 456 GGLPPEGRGMPPPLRGSGRGPRGMSRGRGGRGGGMSAGBGRGPFNK 449  
 DB 406 GGLPPEGRGMPPPLRGSGRGPRGMSRGRGGRGGGMSAGBGRGPFNK 449  
 QY 513 GNVQKAGDMQCPNPGCGNQNFAMRTCCNQCKAPKEGF-LPPFPFGGDRGRGPGGM 571  
 DB 450 GNVQKAGDMQCPNPGCGNQNFAMRTCCNQCKAPKEGF-LPPFPFGGDRGRGPGGM 505



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 16:33:24 ; Search time 7016.69 Seconds  
(without alignments)  
4530.143 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDYSTRSQAAAGCYSA.....GGRGKMKKEHQRDRRY 656

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool.p/US10791017/unal.17022005.125807.22074/app.query.fasta\_1.1358  
-DB=GenBml -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10791017\_@CGN 1\_1.7357 @unal.17022005.125807.22074 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.se.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vl.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3633	100.0	1971 12 BT007796	BT007796 Synthetic
2	3633	100.0	2182 9 BC004817	BC004817 Homo sapi
3	3633	100.0	2371 6 A36460	A36460 Sequence 1
4	3633	100.0	2371 6 AR080100	AR080100 Sequence

5	3633	100.0	2390 6 C0867364	C0867364 Sequence
6	3633	100.0	2390 6 AX411125	AX411125 Sequence
7	3633	100.0	2390 6 HSEWS	X6689 H. sapiens E
8	3618.5	99.6	2164 9 BC072442	BC072442 Homo sapi
9	3618.5	99.6	2189 9 CR456490	CR456490 Homo sapi
10	3618.5	99.6	2364 9 BC011048	BC011048 Homo sapi
11	3605.5	99.2	2189 9 AK056309	AK056309 Homo sapi
12	3602	99.1	2226 6 CQ730352	CQ730352 Sequence
13	3579	98.5	2189 10 BC068226	BC068226 Mus muscu
14	3571.5	98.3	2188 6 AX305537	AX305537 Sequence
15	3571.5	98.3	2188 10 NMEMS	X79233 M. musculus
16	3564.5	98.1	2396 5 BC083960	BC083960 Xenopus l
17	3291.5	90.6	132906 2 AC109802	AC109802 Canis fam
18	3291.5	90.6	163104 2 AC110672	AC110672 Canis fam
19	3277	90.2	1988 6 AX714330	AX714330 Sequence
20	3277	90.2	1988 6 AK056681	AK056681 Homo sapi
21	3233	89.0	193559 2 AC137500	AC137500 Homo sapi
22	3210	88.4	2026 6 C0850483	C0850483 Sequence
23	3210	88.4	2026 6 AK127624	AK127624 Homo sapi
24	3210	88.4	155815 2 AC018774	AC018774 Homo sapi
25	3210	88.4	180718 9 AL596087	AL596087 Human DNA
26	3210	88.4	182501 2 AC011221	AC011221 Homo sapi
27	3209.5	88.3	2440 5 AJ719366	AJ719366 Gallus ga
28	3156	86.9	221444 2 AC113313	AC113313 Mus muscu
29	3100.5	85.3	247757 2 AC121282	AC121282 Mus muscu
30	2822.5	77.7	1783 6 CQ721057	CQ721057 Sequence
31	2819.5	77.6	226400 2 AC128482	AC128482 Rattus no
32	2819.5	77.6	235241 2 AC098231	AC098231 Rattus no
33	2819.5	77.6	263925 2 AC106522	AC106522 Rattus no
34	2730	75.1	7293 9 HSM808920	BS648765 Homo sapi
35	2601	71.6	2665 5 BC063928	BC063928 Xenopus t
36	2586.5	71.2	2683 5 BC075120	BC075120 Xenopus t
37	2338	64.4	270337 2 AC099212	AC099212 Rattus no
38	2292	63.1	221924 2 AC105889	AC105889 Rattus no
39	2292	63.1	275105 2 AC118907	AC118907 Rattus no
40	1908.5	52.5	211189 2 AC128031	AC128031 Rattus no
41	1908.5	52.5	217080 10 AC125565	AC125565 Rattus no
42	1885.5	51.9	2489 5 BC068357	BC068357 Danio rer
43	1880	51.7	2448 5 BC056281	BC056281 Danio rer
44	1816	50.0	2473 9 AK026270	AK026270 Homo sapi
45	1786	49.2	1864 9 AF254086	AF254086 Homo sapi

#### ALIGNMENTS

RESULT 1  
BT007796  
LOCUS  
DEFINITION Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1  
ACCESSION BT007796  
VERSION BT007796.1 GI:30584430  
KEYWORDS FLI CDNA.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 1971)  
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,  
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,  
Pheasant,M. and Farmer,A.  
TITLE Cloning of human full-length CDS in BD Creator(TM) System Donor  
vector  
JOURNAL Unpublished  
REFERENCES 2 (bases 1 to 1971)  
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,  
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,  
Pheasant,M. and Farmer,A.  
AUTHORS Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow  
Circle, Palo Alto, CA 94303, USA  
COMMENT This CDS clone is a part of a collection of human full length  
expression clones generated by BD Biosciences Clontech and the  
Harvard Institute of Proteomics. Each CDS has been cloned in two







Db 20 ATGGCGTCCACGGATTACGACTTACGACGAGCTGACGCGGAGGCGTACAGTGGT 79  
 QY 21 TTTTThAlaGlnProThrGlnGlyTyrAlaGlnThrThGlnAlaTyrGlnGlnSer 40  
 Db 80 TACACGCGCCAGCCACTCAGATATGACACAGCACCCAGGACATATGGCGCAACAAAC 139  
 QY 41 TTTGThThTyrGlnProThrAspValSerTyrThGlnAlaGlnThrThAlaThr 60  
 Db 140 TATGAACTTATGAGAACGCCACTGATGCTATACCCAGGCTCAGACCTGACAC 199  
 QY 61 TTTGThAlaThAlaTyrAlaThrSerTyrGlnProProThrGlnTyrThThPro 80  
 Db 200 TATGGCAGACCGCTATGCACTTCTATGACAGCTCCACTGCTGTTATACTTCA 259  
 QY 81 ThAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlnThThAlaTyrAspThr 100  
 Db 260 ACTGCCCCCAGGACATACAGCCAGCTGTCAGGGGTATGGCACTGGTGTATGATACC 319  
 QY 101 ThThAlaThAlaTh 120  
 Db 320 ACCACTGCTACAGTACACACACCGGCTCTTATGAGCTCAGTCTGATATGGCACT 379  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnProAlaAlaThAlaProThrArgProGln 140  
 Db 380 CAGCTGCTTATCAGCTATGGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 439  
 QY 141 AspGlyAsnIleProThrGlnThThSerGlnProGlnSerThThGlnTyrAsnGln 160  
 Db 440 GATGGAACAAAGCCACTGAGTACATCACTCAATATGACAGGGGGTTTCAACAG 499  
 QY 161 ProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnAlaProGlySerTyrPro 180  
 Db 500 CCAAGCTGAGATATGACAGATTAATCACTTATCCAGGTACTGGAGCTACCC 559  
 QY 181 MetGlnProValThAlaProProSerTyrProProThSerTyrSerSerThThGlnPro 200  
 Db 560 ATGACACAGTCACTGACCTCATCTTACCTTACAGAGTATCTTATCAGAGCG 619  
 QY 201 ThSerTyrAspGlnSerTyrTyrSerGlnGlnThThTyrGlnProProSerTyr 220  
 Db 620 ACTAGTATGATAGAGAGTATCTCTCAGCAGAAACCTATGGGCAACCGAGCAGCTAT 679  
 QY 221 GlyGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnProProThSerTyr 240  
 Db 680 GGAACACAGATGCTATGCTCAACAGAGCTATGGGAGAGCTCCCACTAATTAC 739  
 QY 241 ProProGlnThThTyrSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 740 CCAACCCCAAACTGATCTTACAGCCAAAGCTCAAGTCAATATAGCCAAACAGAGCAGCAG 799  
 QY 261 TTTGThGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlnTyrGlnGln 280  
 Db 800 TAGGGGAGAGAGTTCATTCAGACAGAACCCAGTAGAGTGGTGTATTTATGGGAG 859  
 QY 281 GluSerGlyTyrPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 860 GAGTCTGAGAGATTTCCGACCAAGAGAGAACCCGAGCATGAGTGGCTGTATACCGG 919  
 QY 301 GlyArgGlyArgGlyTyrPheAspArgGlyTyrMetSerArgGlyTyrArgGlyGly 320  
 Db 920 GGCAGGGGAGAGGGGAGTTTGTATCGTGGAGCATGACAGAGTGGGCGGGAGAGAG 979  
 QY 321 ArgGlyTyrMetGlySerIleArgGlyTyrGlnGlyTyrPheAsnIleProGlyTyrPromet 340  
 Db 980 CGGGGTGGAAATGGGAGCGCTGAGAGCGAGTGGCTTCAATAGCCCTGTGTGAGCCATG 1039  
 QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
 Db 1040 GATGAAGAGACCAATCTTGTATGAGCCCACTGTAGATCAATGAAGACTTGTGACAC 1099  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThThLeuAspLeuAlaAspPhe 380  
 Db 1100 AGTGAATTTATGATACAGAGATTAATGACAGTGTGCTTATGATGATGTGCGAGACTTC 1159

QY 381 PheArgGlnCysGlyValValIleMetAsnIleArgThThGlnProMetIleHisIle 400  
 Db 1160 TTTAAGCATGTGGGGTGTATGATGAACAAGAACTGGGCAACCCATGATCCATC 1219  
 QY 401 TTTLeuAspIleGlyThThGlyTyrProIleGlyAspAlaThThValSerTyrGlnAspPro 420  
 Db 1220 TACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGCCCA 1279  
 QY 421 ProThAlaIleValAlaValIleThThPheAspGlyTyrAspPheGlnGlnIleSerIleVal 440  
 Db 1280 CCCACTGCGAAGCTCGCGTGGAAATGTTGATGAGAAAGATTTTCAGAGGAGCAAACTT 1339  
 QY 441 LysValSerLeuAlaArgIleValSerProPheMetAsnSerMetArgGlyTyrLeuProPro 460  
 Db 1340 AAAGTCTCTTGTCTGGAGAAAGCTCCAAATGAACATGATCGGGGTGTCTGCCACCC 1399  
 QY 461 ArgGlnGlyArgGlyTyrMetProProPheLeuArgGlyTyrProGlyTyrProGlyTyrPro 480  
 Db 1400 CGTGAAGGAGAGGATGCCACCACTCGTGAAGGTCCAGAGGCGCCAGAGAGTCTCT 1459  
 QY 481 GlyIleProMetGlyTyrMetGlyTyrArgGlyTyrGlnIleAspArgGlyTyrPheProProArg 500  
 Db 1460 GGGGAGACCATGGTGTGCGATGGAGAGCCGTGAGAGAGATGAGAGAGGCTTCTCCACAGA 1519  
 QY 501 GlyProArgGlySerArgGlyValAsnProSerGlyTyrGlnIleAsnValGlnHisArgAlaGly 520  
 Db 1520 GAACTCCGGGGTCCCGAGAGAACCCCTCTGAGAGAGAAAGCAACCGAGCTGGA 1579  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThThGlnCys 540  
 Db 1580 GACTGGAGATGCCCAATCCGGTGTGGAACCAAGAACTTCCCTGAGAAACAGATGC 1639  
 QY 541 AsnGlnCysIleValAlaProIleProGlyTyrPheLeuProProPheProProProGly 560  
 Db 1640 AACAGGTGTAAAGCCCAAGAGCTTCTCCCGCCACCTTTCCCGCCCGGGGT 1699  
 QY 561 GlyAspArgGlyArgGlyTyrProGlyTyrMetArgGlyTyrArgGlyTyrLeuMetAsp 580  
 Db 1700 GGTATGTTGGAGAGAGTGGCCCTGTGTGATCCGGGAGAGAGAGTGGCTCATGAT 1759  
 QY 581 ArgGlyTyrProGlyTyrMetPheArgGlyTyrArgGlyTyrAspArgGlyTyrPheArg 600  
 Db 1760 CGTGGTGGTCCGGTGAATGTTCAGAGGTGGCCGTGGTGAAGACAGAGTGGCTTCCGT 1819  
 QY 601 GlyIleValArgGlyMetAspArgGlyTyrPheGlyTyrGlnIleArgGlyTyrProGlyTyr 620  
 Db 1820 GGTGGCCGGGAGATGACCGAGGTGGCTTGTGTGAGAGAGACAGGTGGCCCTGGGGGG 1879  
 QY 621 ProProGlyProLeuMetGlnGlnMetGlyTyrArgArgGlyTyrArgGlyTyrProGly 640  
 Db 1880 CCCCTGACCTTTGATGGAACAATGGAGAGAAAGAGAGAGAGCTGAGAGAGCTTGA 1939  
 QY 641 LysMetAspIleGlyTyrHisArgGlnIleArgArgAspArgProTyr 656  
 Db 1940 AAAATGATTAAGGAGAGACCGTCAAGAGGCCAAGATCGGCCCTTAC 1987

RESULT 3  
 A36460  
 LOCUS 2371 bp DNA linear PAT 05-MAR-1997  
 DEFINITION Sequence 1 from Patent WO9323549.  
 ACCESSION A36460  
 VERSION A36460.1 GI:2293778  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2371)  
 Aurias,A., Delattre,O., Desmazes,C., Melot,T., Peter,M.,  
 Plougastel,B., Thomas,G. and Zucman,U.  
 NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
 RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE





[illegible]

DB:	Query Match:	Best Local Similarity:	Mismatches:	Indels:	Gaps:
US-10-791-017A-2 (1-656) x CQ867364 (1-2390)	100.00%	100.00%	0	0	0
QY	1	MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla	20		
DB	44	ATGGGATCCAGGATTTACGATACCTTAAGCCAACTGCAAGCGCAGCGAGGCTTACAGTCT	103		
QY	21	TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer	40		
DB	104	TACACCGCCGACCCCACTCAAGATATAGCAACAACCCAGCGATATGGCAACAAGC	163		
QY	41	TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr	60		
DB	164	TATGGAACCTATGAGACAGCCCACTGATGCTATACCCAGGCTCAGAACCACTGCAACC	223		
QY	61	TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro	80		
DB	224	TATGGGACAGCGGCTATGCAACTCTTATGAGACGCTCCCACTGGTTATACTATCCA	283		
QY	81	ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyTyrThrAlaTyrAspThr	100		
DB	284	ACTGCCCCCGACGATACAGCCACCTCTCCAGGGATATGGCACTGGTGCTTATGATACC	343		
QY	101	ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120		
DB	344	ACCACTGCTACAGTACCAACCAACCGGCTCTCATGCACTGATGCTATATGGCACT	403		
QY	121	GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln	140		
DB	404	CAGCTGCTTATCCAGCTATGGGACAGCAAGCAAGCAAGCACTGCACTCAACAAGCCGAG	463		
QY	141	AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln	160		
DB	464	GATGGAACAACGCCCACTGAGACTATGCTACACTCATCTTACGACAGGGGGTTACAACAG	523		
QY	161	ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180		
DB	524	CCAGCGCTATGATGAGCAGAGATTAACAAGTATATCCCGAGTAACTCGGAGACTAACC	583		
QY	181	MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200		
DB	584	ATGAGACGACGACTGCACTCCATCTCAACCTCTCAACAGCTATCTCTTACACAGCG	643		
QY	201	ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220		
DB	644	ACTGATTTATGATCCAGAGCACTTACTCTCAGAGAAACACTTATGGCAACCAAGCAAGCTAT	703		
QY	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr	240		
DB	704	GGAACGACGAGTACTATGTCACAACAAGCAGCTATGGGACGAGCCTCCCACTAGTTAC	763		
QY	241	ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer	260		
DB	764	CCACCCCAACCTGATCTCTACAGCAAGCTCAAGTCAATATATACCAACAGAGACGAGC	823		
QY	261	TyrGlyGlnGlnSerSerPheArgGlnAspIleProSerSerMetGlyValTyrGlyGln	280		
DB	824	TACGGGACGACGAGTTTCACTTCCAGACAGCAACCCCACTGACATAGGGTGTTATGGGAG	883		
QY	281	GluSerGlyGlyPheSerGlyProGlyGlnAlaGlySerMetSerGlyProAspAsnArg	300		
DB	884	GAGTCTGAGGATTTTCCGACCAAGAGAGAAACCGAGCATGAGTGGCCCTGATACCGG	943		
QY	301	GlyArgGlyValArgGlyGlyPheAspArgGlyGlyMetSerArgGlyValArgGlyGlyGly	320		
DB	944	GGCAGGGGAGAAAGGGGGGATTTGATCTGTGAGACATGAGCAAGTGGCGGGGAGAGAGA	1003		
QY	321	ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMet	340		
DB	1004	CGGGGTGGAATGGGACGCGTGGAGAGGAGGAGGTGGCTTCAATTAAGCTGGTGGACCATG	1063		

QY	341	AspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn	360
Db	1064	GATGAAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGCACAC	11223
QY	361	SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspHe	380
Db	1124	AGTGCATTTATATGATCAAGATTTAAATGACAGTGTGACTTATAGATGATCTGGCAGACTTC	1183
QY	381	PhelybGlnCysGlyValValIleMetLeuAsnIleAspThrGlyGlnProMetIleHisIle	400
Db	1184	TTTAGCAGGTGTGGGTGTGTAAAGTATGAACAAGAGAACATGGGCACCACTGATCCACATC	1243
QY	401	TyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro	420
Db	1244	TACCTGCACAGGAACAGAAAGCCCAAGGCCATGCCACAGTGTCTTATGAGAACCCA	1303
QY	421	ProThrAlaLysValAlaValGluTyrPheAspGlyLysAspPheGlnGlySerIleLeu	440
Db	1304	CCCATGCGCAAGGCTGGCGGTGGAAATGGTTTATGAGGAAGAATTTCAAGGGAGCAAACTT	1363
QY	441	LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro	460
Db	1364	AAAGCTCCTCTGCTCGAGAAAGCCTCCATGAACATATGCGGGGTGTCTGCACCC	1423
QY	461	ArgGluGlyYaaArgGlyMetProProProLeuAspArgGlyGlyProGlyGlyPro	480
Db	1424	CGTAGGGCAGAGGGATGCCACACACACTCTCGTAGAGGTCCAGAGGGCCCAAGAGGTCTT	1483
QY	481	GlyGlyProMetGlyArgMetGlyGlyLysArgGlyGlyLysAspArgGlyGlyPheProProArg	500
Db	1484	GGGGGACCCATGGGTGGCATGGGAAGCGCTGGAGGAAGATAGAGAGGCTTCCCTCCAGAA	1543
QY	501	GlyProAspGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly	520
Db	1544	GGACCCCCGGGTTCCTCGAGGGAAACCCCTCTGAGGAGGAACATGTCACACACCGAGCTGGA	1603
QY	521	AspTyrGlnCysProAsnProGlyLysGlyGlnAsnGlnAsnPheAlaTyrArgThrGluLys	540
Db	1604	GACTGGCAGTGTCCCAATCCGGGTGTGTGAACCAAACTTGTCTGTGAAGAACAGATGCC	1663
QY	541	AsnGlnCysLysValaProLysProGluGlyPheLeuProProPheProProProGly	560
Db	1664	AACCACTTAAAGGCCCCCAAGCTGAAGGCTTCTCCGCCACCTTTCGGCCCCGGGT	1722
QY	561	GlyAspAspArgLysArgGlyGlyProGlyGlyMetAspArgGlyGlyLysGlyGlyLeuMetAsp	580
Db	1724	GGTGATCTGTGGCAGAGTGTGCCCTGTGTGCATGCCGGAGAGAAAGTGTGCTCATGAT	1783
QY	581	ArgGlyGlyLysProGlyGlyMetPheArgGlyGlyLysArgGlyGlyLysAspArgGlyGlyLysArg	600
Db	1784	CGTGTGTGTCCCGGTGAATGTTCAGAGGTGGCCGTGTGTGAGACAGAGTGTGGCTTCGT	1843
QY	601	GlyGlyLysArgGlyMetAspAspArgGlyGlyLysPheGlyGlyGlyLysArgArgGlyGlyLysProGlyGly	620
Db	1844	GGTGGCCGGGGCATGACCGAGGTGGCTTTGTGTGAAGAAACAGAGTGTGGCCCTGGGGG	1903
QY	621	ProProGlyProLeuMetGluGlnMetGlyGlyLysArgArgGlyGlyLysArgGlyGlyLysProGly	640
Db	1904	CCCCCTGCACCTTTATGAGAAACAGATGGGAGGAAGAAAGAGGAGCAGTGGAGGACCTGGA	1963
QY	641	LysMetAspLysGlyGlyLysHisArgGlnGluArgGlyAspArgProTyr	656
Db	1964	AAAAATGATAAAGGCGAGCACCTGTAGAGGCCCAAGATCGGCTTAC	2011

RESULT 6			
AX411125			
LOCUS	AX411125	2390 bp	DNA
DEFINITION	Sequence 3772 from Patent WO0229103.		linear
ACCESSION	AX411125		
VERSION	AX411125.1	GI:21443830	
KEYWORDS			
SOURCE	Homo sapiens (human)		

```

ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE      1
AUTHORS        Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
TITLE          Gene expression profiles in liver cancer
JOURNAL        Patent: WO 0229103-A 3772 11-APR-2002;
                GENE LOGIC INC. (US)

FEATURES       location/Qualifiers
                source
                1. 2390
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                /note="EMBL/Genbank Accession No. X66899"

ORIGIN

```

Alignment Scores:	
Pred. No.:	6,256-77
Score:	3633.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
Ds:	6
US-10-791-017A-2 (1-656) x AY411125 (1-2390)	
	Length: 2390
	Matches: 656
	Conservative: 0
	Mismatches: 0
	Indels: 0
	Gaps: 0

QY	1	MeAlaSerThrAspGlySerThrTyrsSerGlnAlaAlaGlnGlnGlyTyrSerAla	20
Db	44	ATGGCGTCCAGGATTAACGTAACCTTAATACCACTGCAGCGGACGAGGGCTACAGTGT	103
QY	21	TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlyGlnSer	40
Db	104	TACACCGCCGACCGCACTCAAGATATATCAACGACCACCAAGCATATGGGCACAAAGC	163
QY	41	TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr	60
Db	164	TATGGAACTTATGGACAGCCCACTATGCACTTATCCAGGCTCAGACCACTGGCAAC	223
QY	61	TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro	80
Db	224	TATGGGACAGACCGCTATGCAACTCTTATGGACAGCCCTCCACCTGGTATATACTACCA	283
QY	81	ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyTyrThrGlyValaTyrAspThr	100
Db	284	ACTGCGCCCCAGGCAATPACAGCGAGCTGTCCAGGGGTATGGACGTGGCTTATGATACC	343
QY	101	ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
Db	344	ACCACTGGTACAGTACCAACCAACCGAGGCTCTCTATGGACGCTCAGTCTGATATGGCACT	403
QY	121	GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln	140
Db	404	CAGCGCTGTATCCAGCCTATGGGAGGACGACGACCACTGACCTTCACMAACCGCAG	463
QY	141	AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln	160
Db	464	GATGAAACMAAGCCCACTGAGACTGATCACTCAATCTAGCAACAGGGGTTTCAACACAG	523
QY	161	ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	524	CCGAGCCTTAGGATATGGACAGAGTAACTACATTTATCCAGGTATCTGGAGACTACCCC	583
QY	181	MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200
Db	584	ATGACGCCAGTACGACACTCACTCACTACCTCTCAACAGCTATTCTCTTACACAGCCG	643
QY	201	ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220
Db	644	ACTAGTTATGATCAAGACGATTACTCTCAGCAGAACCACTATGGGCACCGACAGACTAT	703
QY	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrThrSerTyr	240
Db	704	GGAACGACAGATGCTATGCTCAACAAAGCACTATGGGACAGAGCTCCCACTAGTTAC	763





polyA\_signal 2162..2167  
 /gene="EWS"  
 /evidence=experimental  
 polyA\_signal 2350..2355  
 /gene="EWS"  
 /evidence=experimental  
 polyA\_site 2371  
 /gene="EWS"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.25e-77 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-791-017a-2 (1-656) x HSEWS (1-2390)

QY	1	MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla	20
DB	44	ATGGCGTCCACGGATTACAGTACCTATAGCCAGCTGCAGCGCAGCGGCTACAGTGC	103
QY	21	TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer	40
DB	104	TACACCGCCCGCCAGCTCAGAGATATGCACAGCCAGCGCATATGGGCAACAAAGC	163
QY	41	TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr	60
DB	164	TATGGAACCTATGAGCAGCGCCATGATGTCAGCTATATCCAGGCTCAGACCTGCAC	223
QY	61	TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro	80
DB	224	TATGGCAGACCGCCATATGCACTTCTATGACAGCTCCACCTGGTATACTACTCCA	283
QY	81	ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr	100
DB	284	ACTGCCCGCCAGCATACAGCCAGCTGTCACGGGATATGGCATGGTGTATGATACC	343
QY	101	ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
DB	344	ACCACGTCTACAGTCAACACACCCAGGCTCTTATGACAGCTCAGTCTGATATGGAC	403
QY	121	GlnProAlaSerProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln	140
DB	404	CAGCCGTCTTATCAGCTATGGGACAGCCAGCCAGCGCCTGACCTTACAAAGCCG	463
QY	141	AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln	160
DB	464	GATGGAAACAGCCCATGAGCTAGTCACTCAATCTAGCACAGGGGTTTCAACAG	523
QY	161	ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
DB	524	CCACAGCTAGATATGACAGAGTAATCACTAGTTATCCCAAGGTACTGGGAGCTACCC	583
QY	181	MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200
DB	584	ATGCAGCGAGTCACTGACCTCACTTACCTTACCCAGCTATTCCTTACACAGCG	643
QY	201	ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnProSerSerTyr	220
DB	644	ACTAGTTATGATGAGGAGTATCTCTGACAGAACCTATGGGCAACCGACACTAT	703
QY	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr	240
DB	704	GGACACACAGGATATGCTCAACAAAGCAGCTATGGGACAGAGCTCCCACTAGTAC	763
QY	241	ProProGlnThrThrLysSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer	260
DB	764	CCACCCCAAACTGATCTTACAGCCAAAGCTCCAAAGTCATATATGCAACAGACACAG	823
QY	261	TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln	280

DB	824	TACGGGACACAGATTCACTCCGACAGACACCCAGTACGATGGGTCTTATATGGCAG	883
QY	281	GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg	300
DB	884	GAGTCTGAGAGATTTTCCCGACACAGAGAAACCCGACCATATGAGCTGCTGATTA	943
QY	301	GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly	320
DB	944	GGCAGGGGAGAGGGGATTTATCGTGGAGCATGAGAGAGTGGCGGGAGGAGGAG	1003
QY	321	ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyPromet	340
DB	1004	CCCGGTGGAATGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCTGCTGGTGA	1063
QY	341	AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn	360
DB	1064	GATGAAGGACACAGATCTTATCTAGGCTCCCTGTATGATCCAGATGACACTTGA	1123
QY	361	SerAlaLeuTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe	380
DB	1124	AGTGCATTTATGTCACAGGATTTAAATGACAGTGTGACTTATGATGATCTGGCAG	1183
QY	381	PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnPrometLleHisIle	400
DB	1184	TTTAAAGCATGTGGGGTGTCTTAAGATGAACAAGAACTGGGCAACCATGATCCAC	1243
QY	401	TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro	420
DB	1244	TACCTGACAAAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAAAGC	1303
QY	421	PromThrAlaLysAlaAlaValGlnThrPheAspGlyLysAspPheGlnGlySerLysLeu	440
DB	1304	CCCACTGCCAAGGCTGCGGTGGAATGGTTTATGAGGAAGATTTTCAAGGACCA	1363
QY	441	LysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyGlyLeuProPro	460
DB	1364	AAAGTCTCCCTTCTCGGAAGAAAGCTCCAAATGAACAGATATGGGGTGTGTG	1423
QY	461	ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro	480
DB	1424	CGTGAAGGACAGGACATGCAACACCACTCCGTGAGGTCACAGAGGCCACAGAG	1483
QY	481	GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyLysPheProArg	500
DB	1484	GGGGACCATATGGTGTGATGGAGAGCCGTGAGAGATAGAGAGGCTTCTCCAA	1543
QY	501	GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly	520
DB	1544	GGACCCCGGGGTTCGAGAGGAACCTCTGAGAGAGGAAGTCCAGCACGAGCTGG	1603
QY	521	AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThrGluCys	540
DB	1604	GACTGGCATGTGCCAATCCGGGTGTGGAACCAAGAACTTGGCTCGGAGAACAG	1663
QY	541	AsnGlnCysValAlaProLysProGlnGlyPheLeuProProProPheProProProGly	560
DB	1664	AACCACTGTAAGGCCCAAGCCCAAGCTTGAAGCTTCTCCGCCACCTTTCCG	1723
QY	561	GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp	580
DB	1724	GGTGAATCGTGGACAGAGTGGCTGTGGTGCATGGCGGAGGAAGAGTGGCTCAT	1783
QY	581	ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyLysPheArgGly	600
DB	1784	CGTGGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGTGGCTTC	1843
QY	601	GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly	620
DB	1844	GGTGGCGGGGACATGGACGAGGTGCTTGTGTGAAGAAACGAGAGTGGCTTGG	1903
QY	621	ProProGlyProLeuMetGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly	640

Db	1904	CCCTCGACCTTATGGAACAGATGGGAGGAAAGAGGAGGACGCTGGAGACCTTGA	1963
Qy	641	LYWETLAPPLYEGLYGLNHIHARGLNGLNLTARGATGAPATGATCTTCTT	656
Db	1964	AAAAAGATATAAGGACGACCGTCATGAGAGCGCAGAGATCGGCTTAC	2011
RESULT #	8		
LOCUS	BC072442		
DEFINITION	2164 bp	mRNA	linear
ACCESSION	BC072442		
VERSION	BC072442.1	GI:48734726	
KEYWORDS	MGC:		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2164) Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Bucio, K.H., Scheaffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Ditchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stepleson, M., Soares, M.B., Donald, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ueda, T.B., Tomshilki, S., Carinini, F., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mulhaly, S.U., Bosak, S.A., McSwain, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wollay, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huly, S.W., Viall, D.K., Muzny, D.M., Sodergren, E., Lu, X., Gibbs, R.A., Paley, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, J.A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, N.C., Rodriguez, A.C., Gilmanow, J., Schmitt, J., Myers, R.M., Buterfield, Y.S., Krzyzanski, M.I., Skalka, U., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26)	16899-16903 (2002)
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2164)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-JUN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a> Tissue Procurement: ATCC/CDT/DTF CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a> Contact: <a href="mailto:nisc.mgc@nih.gov">nisc.mgc@nih.gov</a> Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McQuinn, J., Pearson, R., Shankar, S., Thomas, P.J., Touchman, J.W., Tsurgan, C., Vogt, J.L., Walker, M.A., Wehrdy, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAX Plate: 174 Row: 9 Column: 18 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 4885224.		

```

FEATURES
    source          location/Qualifiers
                    1..2164
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="MGC:87992 IMAGE:6047805"
                    /tissue_type="Skin, melanotic melanoma."
                    /clone_lib="NH_MGC_72"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-Sport6"
gene              1..2164
                  /gene="EMSR1"
                  /note="synonym: EMS"
                  /db_xref="LocustID:2130"
                  /db_xref="MIM:133450"
                  7..1974
                  /gene="EMSR1"
                  /codon_start=1
                  /product="Bovine sarcoma breakpoint region 1, isoform EMS"
                  /protein_id="AAH72442.1"
                  /db_xref="GI:48734727"
                  /db_xref="LocustID:2130"
                  /db_xref="MIM:133450"
                  /translation="MASTDYSTYSQAALAQGYSAVTAPPTGVAQTTOAVVGQSYSGT
COPTRVSTTGAQTATTGTGGTAATATSGDPPETGTTPTPAQAASQPPIQGCTGTAVDITTT
AIVTTQASLYAAQSAYVGTQAPAYAGQQQAALAPRPDUQNKPTESISQPSSTGTNTN
PSLGGQSNYSYPVOGVSPMPQVTPAIPESPPTSYSSTQPTSDSSVSQONTYPQGP
SYGOOSSYGQOOSYSCQOPTSYLPQTGSYQAAPSQOOSSEPDODHPISMSP
VYSGOOSYSGFGENRSMMSGPNNRGGRGDPFRGGKSRGGRRGMAGGERGFNR
PGCBWDEGPDLDLGPVPDDEDSDNSAIYVGLINDSNVTLDDLADPFKKQSVVMNRKR
GQPMHIHYLDKETGPKEDATVSYEDPPTAKAAVEPDFSKDROGSLTKGISLARXKRP
NSMRGJLPPREBERGMPPLIRGPGPGGPGGWGNRWGRGSDRGFPPEPRSRKNG
SGGAMVORHARGADMQCNPDCGQNPNAMRTENCNQCAKPKEGFLPEPPPGDGRRGKG
PCGMWRGEGJLMDRGAPGCPMGFCRGGRGDRGFGRRGGRGMDRGFGCGRRRGCGP
MEQWGRGRGRGRCGPGKMKDKGEHRGRBRDPY"
ORIGIN
Alignment Scores:
Pred. No.:      1..28e-76      Length:      2164
Score:          3618.50        Matches:     655
Percent Similarity: 99.85%      Conservative: 0
Best Local Similarity: 99.85%   Mismatches:  0
Query Match:    99.60%         Indels:      1
DB:             9              Gaps:          1

US-10-791-017A-2 (1-656) x BC072442 (1-2164)
QY      1 MetAlasertHAsptTyserThrTYrsSerGlnAlaAlaGlnGlnGlyTYrsSerAla 20
Db      7 ATGGGGTCACCGAATTAAGTACTATACCCAAAGCGACGCCAGCGAGGAGCTTAACAATGCT 66
QY      21 TYrThrAlaGlnProThrgInglYTyraGlnInthrThgAlaAlaTyrgIyngInser 40
Db      67 TACACC GCCCACCCACCTCAAGAGATATGACAGAACCCACCGAGCATATGGGCAACAAAGC 128
QY      41 TyrgIyThrTYrgIyGlnProThrxAspValaserTYrThGlnAlaGlnInthrThralathr 60
Db      127 TATGGAACCTATGAGACAGCCACTGATGTCACTATACCCAGGCTCAAGCACTGCAACC 186
QY      61 TyrgIyGlnInthrAlaTYrAlaThrseryTYrgIyGlnProProThrgIyTYrThrPro 80
Db      187 TATGGGACAGACCGCTATGCACTTCTTATGACAGACCCCACTGGTTATACTACTCCA 246
QY      81 ThrAlaProGlnAlaTYrsSerGlnProValGlnGlyTYrGlyThrGlyAlaTYrAspThr 100
Db      247 ACTGCCCCCCAGGACATACAGCAAGCCTGTCCAGGGGTATAGGCACTGTGCTTATGATAACC 306
QY      101 ThrThralathrValThrThrThrgInalaserTYrAlaAlaGlnSeralatYrtyrGlyThr 120
Db      307 ACCATGCTGACAGTCAACCAACCCAGGCTCTCTATGACGCTCAGTCTGATATGCAACT 366
QY      121 GlnProAlaTYrProAlaTYrgIyngInlProAlaThrThralaProThzArgProGln 140

```







[illegible]

OY		501	GlyProArgGlySeArgGlyVAspProSerGlyIyGlyVAmValAlGlnHtArxAlacLy	520	
Db		1525	GGACCCCGGGGTTCGCCAGGGAAACCCTTCTGAGAGGAAGAACCCTCAGCACCAACTGCA	1588	
OY		521	AspITpGInCyProAsnProGlyICysGlyVAnGlnAanPhealAtPrArgThrClucys	540	
Db		1585	GACTGGCAGTGCTCCCAATCCCGGTTGTGGAAACCAAGAACCTTGCGCTGGAGAAACGAGTGCC	1644	
OY		541	AnGInCyVbValAProLYsPProGluGlyPheLeuProPProPProGly	560	
Db		1645	AACAAGTGTAAGGGCCCCCAAGCCTGAAGGCTTCCTCCGCCACACCTTCCCCTCCCGGGGT	1704	
OY		561	GlyAspArgGlyYARgGlyVGIYProGlyIGlyWMeArGlyVGIYARgGlyVGIYLeuMetAsp	580	
Db		1705	GGTATGTCGTGCACAGAGGGGCCCTGTGGCATGCGGGAGGAAGAAGGGTGGCTCATGGAT	1764	
OY		581	ArgGlyVGIYProGlyVGIYMeLPheArGlyVGIYARgGlyVGIYAspArgGlyVGIYPheArg	600	
Db		1765	CGTGGTGGTCCCGGTGGAATTGTCAAGAGTGCGCGTGGTGGAGACAGAGTGCGCTTCCGT	1824	
OY		601	GlyIyYARgGlyWMeAspArgGlyVGIYPheGlyGlyVGIYARgArgGlyVGIYProGlyVGIY	620	
Db		1825	GGTGGCCGGGCGACAGACCGAGGTGGCTTGTGGAGGAAGACGAGGTGGCCCTGGGGGGG	1884	
OY		621	ProProGlyYProLeuMetGluGlnMerGlyVGIYArGrArgGlyVGIYARgGlyVGIYProGly	640	
Db		1885	CCCCCTGGACCTTTGATGAGGAACAATGATGGAGAAAGAAAGAGAGGACGTGGAGGACCTGGA	1944	
OY		641	LYsMeLAspLYsGlyVGIYLIHtEArgGlyIGluYrgrArgAspArgProLYR	656	
Db		1945	AAAATGATTAAGCGCAGCACCGTCGAGAGCGCGAGAGTGGCCCTTAC	1992	
RESULT 11					
AKO56309					
LOCUS					
DEFINITION	Homo sapiens cDNA FLJ13747 fis, clone NTNR12007377, highly similar to RNA-BINDING PROTEIN EWS.	2189 bp	mRNA	linear PRI 30-JAN-2004	
ACCESSION	AKO56309.1				
VERSION	AKO56309.1 GI:16551673				
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS	Oae,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,F., Ishii,S., Yamamoto,J., Saito,K., Kanai,Y., Isoino,Y., Nakamura,Y., Nagaharui,K., Murakami,K., Yasuda,T., Iwayanagi,T., WagaTsuna,M., Shiratori,A., Sudo,H., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yanazaki,M., Nitomoya,K., Ishibaehi,T., Yamashta,H., Hiraoka,K., Fujimori,K., Tanai,H., Kinata,M., Watanabe,M., Hirakawa,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yoshida,M., Hotaru,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A., Sasaki,N., Aocueka,S., YoshiKawa,Y., Matsunawa,H., Ichihara,T., Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hisligeki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Osaki,K., Hiroe,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchiu,S., Itoh,T., Shigetaki,K., Senba,T., Matsunura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togeshi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Cyran,T., Shirai,Y., Takenashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuno,Y., Yamashta,R.,				





Db 1099 GATGAAGACTGTGCAACAGTGCATTTATGTAACAAGATTAATGACAGTGTGACTCTA 1158  
 Qy 375 AAPPAPLeuAlAaPhePheLeuGlnCyGlyValValIyMeCaAnlyBaTgThrGly 394  
 Db 1159 GATGATCTGGCAGACTCTTTAAAGCAGTGTGGGTGTGTAAGATGAACAAGAACCTGGG 1218  
 Qy 395 GlnProMetIleHisIleTyrLeuAaPlyGluThGlyLysProIyGlyBaPAlaTtr 414  
 Db 1219 CAACCATATATCCATCTACTCTGACAGAAACAGAAAGCCAAAGGCGATGCGACA 1278  
 Qy 415 ValSerTyrGluAaPProProThraIaIyBaIaIaValGluTrrPheAaPglyLysAaP 434  
 Db 1279 GTGTCTTATGAAGACCCACCACTGCCAAGGCTGGTGAATGTTTATGGGAAAGAT 1338  
 Qy 435 PheGlnGlySerIyLeuLysValSerLeuAlaArgLysIyPProPheCaAnSerMet 454  
 Db 1339 TTTCAAGGAGAGCAAACTTAAGCTCTCTTGCTCGAAGAACCTCAATGAACAGTATG 1398  
 Qy 455 ATGGGlyGlyLeuProProAArgIyAArgGlyLysProProLeuAaPglyIyGlyPro 474  
 Db 1399 CGGGTGTGTCTGACCCCGTGAAGGAGGAGCATGCAACCACTCCGTGGAGGTCCA 1458  
 Qy 475 GlyIyLysProGlyIyLysProGlyIyProMetGlyAArgMetGlyIyAArgGlyAaPArg 494  
 Db 1459 GGAAGGCCAGAGAGTCTCTGGGGAGCCCATGGGTGCGATGGAGGCGCTGGAGAGATGA 1518  
 Qy 495 GlyIyLysProProAArgIyLysProAArgIySerAArgGlyAaPProSerGlyIyGlyAaP 514  
 Db 1519 GGAAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCTCGAGGGAACCTCTGGAGAGAGAAAC 1578  
 Qy 515 ValGlnHisArgAlaGlyAaPTrpGlnCyAProAaPProGlyIyCyGlyIyAaPProIyAaPPro 534  
 Db 1579 GTCCAGCAGCAGAGCTGAGAGCTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTC 1638  
 Qy 535 AlaTyrArgThrGlnCyAaAngInCyAlyBaIaPProIyAProGlyIyPheLeuProPro 554  
 Db 1639 GCCTGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGGCTTGAAGGCTTCCCGGCA 1698  
 Qy 555 ProPheProProProGlyIyAaPArgGlyIyAArgGlyIyGlyProGlyIyLysMetAaPArgIyGly 574  
 Db 1699 CCTTTCCTCCCGCCGGGGTGTATCGTGGCAGAGTGGCCCTGTGGCATGCGGGAGAGA 1758  
 Qy 575 AArgGlyIyLysMetAaPArgGlyIyLysProGlyIyMetPheAaPArgIyGlyAaPArgIyGly 594  
 Db 1759 AAGAGTGGCTCATGATGTGTGTGTCCCGGTGAATGTCAGAGTGGCTGTGTGA 1818  
 Qy 595 AaPArgGlyIyLysPheAaPArgIyAArgGlyLysMetAaPArgGlyIyGlyPheGlyIyGlyAaPArg 614  
 Db 1819 GACAGAGGTGGCTTCGTGTGTGGCCGGGCGATGGAACGAGGTGGCTTTGTGTGAAGAGA 1878  
 Qy 615 AArgGlyIyLysProGlyIyLysProProGlyIyProLeuMetGlnGlnMetGlyIyLysAaPArgIy 634  
 Db 1879 CAGAGTGGCTCTGGGGGGCCCTGTGACCTTTGATGAAACAGATGGAGAGAAAGAGA 1938  
 Qy 635 GlyAArgGlyIyLysProGlyIyLysMetAaPArgGlyIyLysHisArgGlnIyLysAaPArg 654  
 Db 1939 GGAAGGTGGAGAGCTGGAATAATGATAAAGGAGACCGTCAAGAGGCGAGAGATCGG 1998  
 Qy 655 ProTyr 656  
 Db 1999 CCTTAC 2004  
 RESULT 12  
 CQ730352  
 LOCUS CQ730352 2326 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 16286 from Patent WO02068579.  
 ACCESSION CQ730352  
 VERSION CQ730352.1 GI:42303963  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
 TITLE Kite, such as nucleic acid arrays, comprising a majority of  
 humenexons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;  
 PE Corporation (NY) (US)  
 FEATURES  
 source Location/Qualifiers  
 1..2326  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,22e-76 Length: 2326  
 Score: 3602.00 Matches: 652  
 Percent Similarity: 99.85 Conservative: 0  
 Best Local Similarity: 99.85 Mismatches: 0  
 Query Match: 99.15 Indels: 1  
 DB: Gaps: 0  
 US-10-791-017A-2 (1-656) x CQ730352 (1-2326)  
 Qy 5 AAPPYrSerThrTyrSerGlnAlaAlaGlnGlnIyTyrSerAlaTrrThraIaGln 24  
 Db 8 GATTACAGTACTTATGACCAAGCTGAGCGAGAGAGGCTTACAGTGTACACCGGCCAG 67  
 Qy 25 ProThGlnGlyTyrAlaGlnThrThraIaIyGlyGlnGlnSerTyrGlyTrrTyr 44  
 Db 68 CCACTCAAGATATGCAAGACCAAGCCAGGATATGGGCAACAAAGCTATGGAACCTAT 127  
 Qy 45 GlyIyLysProThraPlyAaPlySerTyrThraIaGlnThrThraIaTrrTyrGlyIyLys 64  
 Db 128 GGAAGCCCACTGATGTGACTATACCCAGGCTCAGACCACTGCAACCTATGGGAGAGCC 187  
 Qy 65 AlaTyrAlaIleThrSerTyrGlyIyLysProProThraIyTrrThrProThraIaPProGln 84  
 Db 188 GCTATGACACTTCTTATGAGACAGCTCCACCTGTTATATCTCACTGACCGCCCAAG 247  
 Qy 85 AlaTyrSerGlnProValGlnIyTyrGlyTrrGlyAlaTyrAaPThrThraIaTrr 104  
 Db 248 GCATACAGCAGACCTGTCCAGGGGTATGGCACTGGGTCTTATGATACCACTGCTTAC 307  
 Qy 105 ValThrThrThraIaIleSerTyrAlaAlaGlnSerAlaTyrGlyTrrGlnProAlaTyr 124  
 Db 308 GTACACCAACCAAGGCTCTTATGACGCTCAGTGTGATATGGCACTGACCTGCTTAT 367  
 Qy 125 ProAlaTyrGlyIyLysProIaAlaIleThrAlaPProThraIyPProGlnIaPProGlyIyAaPly 144  
 Db 368 CAGGCTATGGGAGAGAGCAGCAGCACTGCACTTCAAGACCGAGAGTGAACAG 427  
 Qy 145 ProThraIyThrSerGlnProGlnSerSerThrGlyIyTyrAaPProIyLysGly 164  
 Db 428 CCACTGAGACTATGTAACCTTATATGACAGAGGGGTTCACACAGCCAGCTTACGA 487  
 Qy 165 TyrGlyIyLysSerAaPlySerTyrProGlnValProGlyIySerTyrProMetGlnProVal 184  
 Db 488 TATGACAGAGTATACAGATTATCCCAAGTACCTGAGAGCTACCCCATGAGCAAGTC 547  
 Qy 185 ThrAlaPProProSerTyrPProProThraIySerSerThrGlnPProThraIyAaP 204  
 Db 548 ACTGACCTTCATCTTACCTTCAACAGCTATCTCTTACACAGCGCACTAGTATGAT 607  
 Qy 205 GlnSerSerTyrSerGlnIaPProThraIyGlnIyLysProSerSerTyrGlyIyGlnIyLys 224  
 Db 608 CAGAGAGTATCTCTCAGAGAAACACTTATGGGCAACCAAGAGCTATGAGACAGCAAGT 667  
 Qy 225 SerTyrGlyIyLysSerSerTyrGlyIyGlnIyLysProProThraIySerTyrPProGlnIy 244  
 Db 668 AGCTATGTGCAACAAAGAGCTATGGGAGAGGCTCCCACTATGTTACCAACCCCAAGCT 727  
 Qy 245 GlySerTyrSerGlnAlaPProSerGlnIySerGlnIyLysSerSerTyrGlyIyGlnIy 264

```

Db      728 GGAATCTCAAGCAGAGCTCCAGTCAATATAGCCAAAGAGAGAGCTACGGGAGCG 787
Qy      265 SerSerPheArgGlnAphIspSerSerMetGlyValTyrGlyGlnGlnSerGlyGly 284
Db      788 AGTTCAATCCGACAGAGACCACTCCAGTACATGGGTGTTATGGGACAGAGCTCGAGGA 847
Qy      285 PheSerGlyProGlyGlnAphAaspSerMetSerGlyProAaspAaspArgGlyValArg 304
Db      848 TTTTCCGAGCCAGAGAGAAACCGAGACATGAGTGGCTCGATTAACCGGAGCGAGGAGAA 907
Qy      305 GlyGlyPheAaspArgGlyGlyMetSerArgGlyGlyValArgGlyGlyValMet 324
Db      908 GGGGGATTGATCTGTGAGGAGCAGACAGAGTGGGCGGAGAGAGAGAGCGGTGGAAG 967
Qy      325 GlySerAlaGlyGlnArgGlyGlyPheAaspProGlyGlyProMetAaspGlyPro 344
Db      968 GCGAGCGCTGAGAGAGAGAGTGGCTTCAATTAAGCTGTGAGACCATGATGAGAGACCA 1027
Qy      345 AapLeuAapLeuGlyProProValAapProAaspGlyAaspSerAaspAaspAasp 364
Db      1028 GATCTTGATCTAGGCTCCAGCTGAGATCCAGATGAGAGAGCTGAGACAGTGCATTTAT 1087
Qy      365 ValGlnGlyLeuAaspAaspSerValThrLeuAaspAaspLeuAlaAaspPheGlyGln 384
Db      1088 GTTCAAGATTAATGACAGTGTGATCTGATGATGATCTGGCAGACTTCTTTAAGCAGTGT 1147
Qy      385 GlyValValIleMetAaspAaspAaspArgGlyGlnProMetGlnIleSerIleLeuAasp 404
Db      1148 GGGGTGTATTAAGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1207
Qy      405 GlnThrGlyValProGlyGlyAaspAlaThrValSerGlyGlnAaspProProThrAla 424
Db      1208 GAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
Qy      425 AlaAlaValGlnTrpPheAaspGlyGlyAaspPheGlnGlySerIleLeuValSerLeu 444
Db      1268 GCTGCGGTGGAAGGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1327
Qy      445 AlaArgGlyValProProMetAaspSerMetArgGlyGlyLeuProProArgGlyValArg 464
Db      1328 GCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1387
Qy      465 GlyMetProProProLeuAaspArgGlyGlyProGlyGlyProGlyGlyProMet 484
Db      1388 GGCATCCACACACACTCCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1447
Qy      485 GlyArgMetGlyValArgGlyGlyAaspArgGlyGlyPheProProArgGlyProArgGly 504
Db      1448 GGTGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1507
Qy      505 SerArgGlyAaspProSerGlyGlyGlyValAaspValGlnIleAaspArgGlyValArg 524
Db      1508 TCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1567
Qy      525 ProAaspProGlyGlyValAaspGlnAaspPheAlaTrpArgThrGlnGlnAaspGlnGln 544
Db      1568 CCCAATCCGGGTGTGAGAAACAGAACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627
Qy      545 AlaProLyaspProGlyGlyPheLeuProProProPheProProProGlyGlyAaspArgGly 564
Db      1628 GCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1687
Qy      565 ArgGlyGlyProGlyGlyMetAaspArgGlyGlyValArgGlyGlyLeuMetAaspArgGly 584
Db      1688 AGAGGTGGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1747
Qy      585 GlyGlyMetPheArgGlyGlyValArgGlyGlyAaspArgGlyGlyPheArgGlyValArg 604
Db      1748 GGTGAGATGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1807
Qy      604 ymetAaspArgGlyGlyPheGlyGlyValArgArgGlyGlyProGlyGlyProProGlyPro 624

```

```

Db      1808 CATGACAGAGGTGCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1867
Qy      624 OleMetGlnGlnMetGlyValArgArgGlyGlyValArgGlyGlyProGlyGlyMetAasp 644
Db      1868 TTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1927
Qy      644 sglYglnHsArgGlnGlnArgArgAaspArgProTyr 656
Db      1928 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964

RESULT 13
BC068226
LOCUS
DEFINITION
Mus musculus Bwing sarcoma breakpoint region 1, mRNA (CDNA clone
MGC:61279 IMAGE:5707017), complete cds.
ACCESSION
BC068226
VERSION
BC068226.1 GI:46399228
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2189)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Scheffer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McManus,P.J.,
McKernan,K.J., Malek,J.A., Gamarale,P.H., Richards,S.,
Wortley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W.,
Villalón,D.K., Wuzny,D.M., Sodergren,B.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Keltman,M., Madan,A.C., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitt,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalka,U., Smaliv,D.B.,
Scherer,A., Schein,J.E., Jones,S.J. and Watta,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2189)
Director MGC Project.
Direct Submission
Submitted (30-MAR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaabs-rt@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Schetz,T., Smith,C., Smith,E., Tack,D., Trout,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNU at: http://image.lnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6679714.
Location/Qualifiers
FEATURES

```







```

Db 477 GATGTAACAAGCTGTGAGCTAGCAACCTCAATCTAGCAAGGGGTTATTAACCA 536
Qy 161 ProSerLeuGlyTyrGlyGlnSerAmetTyrSerTyrProGlnValProGlySerTyrPro 180
Db 537 CCCAGCTAGATATGATGACAGATTAACCTACAGCTATCCCAAGGTACCTGGAGCTTACCA 596
Qy 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db 597 ATGAGCAAGTACACCGACCTTCACTTATCTCTTACAGCTACTCTCTTCAACAGCG 656
Qy 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220
Db 657 ACTAGTTACATCAGAGCACTTACTCTCAGCAGAACCTATGGGAGCGGACAGACTAT 716
Qy 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db 717 GGAACAAGAGTATGATGTCACAAAGCAAGCTATGGGAGCAGCTCTCTACTAGTTAC 776
Qy 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260
Db 777 CCGCTTCAAGCTGATCTTACAGCCAGGCTTCCAGTCAATATATGCCAAACAGACAGAC 836
Qy 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
Db 837 TAGGGGAGAGGATTCATCCGACAGACCAACCCAGTAGCAGTGGTTTATGGGAG 896
Qy 281 GluSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerMetSerGlyProAspAsnArg 300
Db 897 GAGCTGGAGGATTTTCCGAGCAGGAGAACCCGAGCTTGAAGTGGCCCTGATTAACCG 956
Qy 301 GlyArgGlyValArgGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320
Db 957 GGCAGGGGAGAGGGGAGTTTGAATCGTGGAGCATGAGAGAGTGGCCGGGAGAGAG 1016
Qy 321 ArgGlyGlyMetGlySerAlaGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet 340
Db 1017 CGCGGTGAGCTGGGC---GCTGGAGAGCGAGGTGGCTTCAATTAAGCTGGTGAACCATG 1073
Qy 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAsn 360
Db 1074 GATGAAGAGCAACATCTTGAATCTTAAGGCTTCTTAATATCCCATGAAGACTCTGAAC 1133
Qy 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380
Db 1134 AGTGCATTTATGTGCAAGATTAAGACATGTGACTGTGATGATCTGGCACTTC 1193
Qy 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
Db 1194 TTTAAGCAGTGTGGGTTGTCAAGATGAACAAGAACTGACAAACCATGATCCATATC 1253
Qy 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420
Db 1254 TACCTGATTAAGGAGACAGAAAGCTTAAGGGGAGCCCAAGTGTCTTAAGAAGATCCA 1313
Qy 421 ProThrAlaLysAlaValAlaValGlyLysPheAspGlyLysAspPheGlnGlySerLysLeu 440
Db 1314 CCAACTCAAAAGCTGCGTGGAAATGTTGATGGGAAAGATTTTCAAGAAAGCAAACTT 1373
Qy 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProPro 460
Db 1374 AAAGTGTCTTGGCCGAAAGAGCTTCAATGAACAAGACGCGGGAGGAGCATGCCACT 1433
Qy 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480
Db 1434 CGTGAAGGAGAGGGGTATGCAACCACTTGTGAGAGTCTGTGTGGCCCAAGAGGCGCTT 1493
Qy 481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyValAspArgGlyGlyPheProProArg 500
Db 1494 GAGAGACCCATGGGTCCATGGAGGCGCTGGAGAGACAGAGGGGAGCTTCCCTCCAAAGA 1553
Qy 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyValAsnValGlnHisArgAlaGly 520

```

```

Db 1554 GGGCCCGAGGCTCCAGAGAAACCTCTGGAGGAGGAAATGTCCAGCAACCGAGCTGA 1613
Qy 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaThrArgThrGlyCys 540
Db 1614 GACTGGCAGTGTCCCAATCCGGGCTGTGAAACCAAGAACTTGTGGAGAAACAGATGC 1673
Qy 541 AsnGlnCysLysAlaProLysProGlyGlyPheLeuProProProPheProProGly 560
Db 1674 AACCAATGAAGGCCCTTAAGCCGAGGGCTTCTCCGCGCAACCTTTCACCTCGGGT 1733
Qy 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyValArgGlyGlyLysMetAsp 580
Db 1734 GGTGATCGTGAAGAGGTGGCCCTGTGGCATGCAAGAGGAAAGAGAGACTCATGAGC 1793
Qy 581 ArgGlyGlyProGlyGlyMetPheArgGlyValArgGlyGlyValAspArgGlyGlyPheArg 600
Db 1794 CGTGGTGTCTTGAGAAATGTTCAAGGTGGCAAGGTGGAGACAGAGAGGCTTCCGA 1853
Qy 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620
Db 1854 GGTGGCCGTGAATGAGCCGAGGTGGCTTGTGGAGAGAAAGAGTGTCTCGGGGGG 1913
Qy 621 ProProGlyProLeuMetGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640
Db 1914 CCTCTGGAACCTTATATGAAACAGATGGGAGAGAGAGGCGGACGTGAGAGACCTGGG 1973
Qy 641 LysMetAspLysGlyGlyHisArgGlnGlyArgArgAspArgProTyr 656
Db 1974 AAAATGATTAAGGAGGAGCAACCTGACAGAACCAAGAGCCGCTTAC 2021

```

Search completed: February 21, 2005, 02:07:17  
 Job time : 7076.69 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 16:18:03 ; Search time 830.229 Seconds  
(without alignments)  
4677.443 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTVSTYSQAQAAGQYGA.....GGPKMDKSHRQERDRPY 656

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlp  
-Q/cg2.1/USPTO.spool.p/US10791017/runat.17022005.125806.22064/app.query.fasta\_1.1358  
-DB=Geneseq.16Dec04 -QPM=fastap -SUPFX=p2n.rng -MINMATCH=0.1 -IOOPL=0  
-IOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=epc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10791017 @CGN 1.1 886 @runat.17022005.125806.22064 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq.16Dec04.\*

1: geneseqn1808.\*  
2: geneseqn1908.\*  
3: geneseqn2008.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002as.\*  
7: geneseqn2002bs.\*  
8: geneseqn2003as.\*  
9: geneseqn2003bs.\*  
10: geneseqn2003cs.\*  
11: geneseqn2003as.\*  
12: geneseqn2004as.\*  
13: geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2371	2	AAQ50643
2	3633	100.0	2372	13	ADQ86032
3	3633	100.0	2390	6	ABK84628 Human tum
4	3633	100.0	2390	6	ABK64822 Human CDN
5	3633	100.0	2390	6	ABN97274 Human ben
					Abn97274 Gene-#377

6	3633	100.0	2390	13	AD816277 Human CDN
7	3633	100.0	2330	13	ACN40903
8	3605	99.2	2176	6	AA862262
9	3571.5	98.3	2188	6	AB199383 Mouse isc
10	3531.5	97.2	2388	5	AA870647
11	3483	95.9	2273	6	AA862623 DNA encod
12	3315	91.2	3309	2	AAQ50646
13	3277	90.2	1988	10	AD853506
14	3210	88.4	2026	13	AD807446
15	3210	88.4	2177	13	ACN37537
16	3178.5	87.5	1807	13	ADP56333
17	1444	39.7	1431	6	AB873253
18	1420	39.1	2403	2	AA87870
19	1420	39.1	2412	2	AA87869
20	1385	38.1	1462	4	AA87869
21	1378	37.9	856	13	ACN37540
22	1323.5	36.4	845	13	ACN37538
23	1290	35.5	1832	12	AD126112
24	1282	35.3	1824	6	AB873272
25	1282	35.3	1824	6	ABK84149
26	1282	35.3	1824	12	AD126116
27	1282	35.3	1824	12	ADU62788
28	1282	35.3	1824	13	ACN39977
29	1282	35.3	2179	4	AAK51488
30	1281.5	35.3	1822	6	AB873275
31	1281.5	35.3	1822	13	AD814648
32	1281	35.3	2209	4	AAK51489
33	1273.5	35.1	1939	12	AD877157
34	1249	34.4	2213	4	AAK52472
35	1249	34.4	2213	4	AAK52473
36	1181	32.5	2299	11	ACN88697
37	1046.5	28.8	1610	11	AD12631
38	1010	27.8	2153	6	ABK84369
39	1010	27.8	2153	12	AD126208
40	1006	27.7	2144	12	AD126208
41	1003.5	27.6	2126	3	AA864661
42	961	26.5	550	9	ACH40478
43	950	26.1	684	13	ACN37539
44	868.5	23.9	568	9	ACH22345
	839	23.1	540	4	AA117281

# ALIGNMENTS

RESULT 1	AAQ50643	standard; cDNA; 2371 BP.
ID	AAQ50643	
XX	AAQ50643;	
AC	AAQ50643;	
XX		
DT	25-MAR-2003	(revised)
DT	26-MAY-1994	(first entry)
XX		
DE	Human Ews gene B1AC5 from foetal brain cDNA library.	
XX		
KW	chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;	
KW	malignant melanoma; hum-flt-1;	
KW	primitive peripheral neuroectodermal tumour; human chromosome 11;	
KW	human chromosome 22; ss.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	25..1995	
FT	CDS	
FT	/tag= a	
FT	/product= "EWS_protein"	
FT	/transl_except= pos:1729..1731; aa:Val	
FT	polyA_signal	
FT	2143..2148	
FT	polyA_signal	
FT	2331..2336	
FT	/tag= b	
FT	polyA_signal	
FT	2331..2336	
FT	/tag= c	
XX		
PN	MO9323549-A2.	



XX 25-NOV-1993.  
 PD 19-MAY-1993; 93MO-FR000494.  
 XX 20-MAY-1992; 92FR-00006123.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougasael B,  
 PI Thomas G, Zucman J;  
 DR WPI; 1993-386580/48.  
 XX P-PSDB; AAR4555.  
 XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion  
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
 XX Disclousure; Fig 6; 123pp; French.  
 XX The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA  
 CC library (Stratagene cat. # 936206). The clone BFLACS was identified and  
 CC sequenced. It represents the entire coding region and 3'-UTR of the Ews  
 CC gene. (Updated on 25-MAR-2003 to correct PN field.)  
 XX Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Prod. No.: 1.69e-116 Length: 2371  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-791-017a-2 (1-656) x AAC50643 (1-2371)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 25 ATGGCGTCCAGGATTACAGTACCTTATGCGAAGCTGACGAGCGAGGCTTACAGTCT 84  
 QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 40  
 DB 85 TACACCGCCGAGCCCACTCAAGATATGACACACCCAGGCAThrGGGCAACAAAGC 144  
 QY 41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60  
 DB 145 TATGGAACCTATGAGACGCCCACTGATGTCACTATACCCAGGCTCAGACCTGCAACC 204  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80  
 DB 205 TATGGGACAGCCGCTATGCACTTCTTATGACAGGCTCCACGTGTATTACTACTCCA 264  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100  
 DB 265 ACAGCCGCCGAGGCAATACAGCCAGCCCTGTCCAGGGGATGGCACTGTGCTTATGATACC 324  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 DB 325 ACCACGTCTACAGTCAACACACCCAGGCTCTTATGACAGTCACTGCAATGACACT 384  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln 140  
 DB 385 CAGCCGCTTATCAGCTATGAGGAGCAGCAGCCAGCCACCTCACCTCAACACCGCAG 444  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160  
 DB 445 GATGAAACAGCCCACTGAGTCACTCACTCACTTACGACAGGGGTTTCAACCG 504  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 505 CCACAGCTAGATATGAGACAGATTAACCTTATCCAGGTAACCTGGAGACTACCC 564

QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 565 ATGACAGCCAGTCACTGACCTCATCTTACCTTACCAAGCACTATTCCTTACACAGCCG 624  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnProSerSerTyr 220  
 DB 625 ACTAGTTATGATCAAGAGCACTTACTTCAAGAGAACTATAGGCAACCGACAGTAT 684  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 DB 685 GACACAGAGATGATGATGTCACAAACACCTATGAGCAGCAGCCCTCCACTAGTTAC 744  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 DB 745 CCACCCCAACTGAGATCTTACAGCCAGGCTCCAGTCAATATAGCCAAACAGACAGCAGC 804  
 QY 261 TyrGlyGlnGlnSerSerPheAlaGlnAlaProSerSerMetGlyValTyrGlyGln 280  
 DB 805 TACGGGCAACAGATTCATTCGACAGACACCCACATAGCATGGGTGTTATGGGCA 864  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAlaGlnSerSerMetGlyProAspAsnArg 300  
 DB 865 GAGTCTGAGAGATTTTCCGAGACAGAGAAACCGAGCATGATGGCCCTGATTAACCG 924  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 DB 925 GCGAGGAGAGAGAGGAGATTTATCGTGGAGCATGAGAGAGTGGCGGGAGAGAGA 984  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 DB 985 CGCGGTGGAATGGGAGCGGCTGAGAGCGAGGTGCTTCAATAGCTGTGGTGAACCATG 1044  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360  
 DB 1045 GATGAAGGACAGATCTTATGATCAGCCCTCTGTATGATCCAGATCAAGCTCTGCAAC 1104  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1105 AGTCAATTTATGATCAAGATTTAAATGACAGTGTGACTTATGATATCTTGCAACCTTC 1164  
 QY 381 PheLysGlnCyseGlyValValLysMetAsnLysArgThrGlyGlnProMetIleAspIle 400  
 DB 1165 TTTAAGCAATGGGGGTTTAAAGATGAACAGAAAGAACTGGGCAACCATGATCCACATC 1224  
 QY 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyAspPro 420  
 DB 1225 TACCTGACAAAGGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAAGCCA 1284  
 QY 421 ProThrAlaLysAlaAlaValGlnTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1285 CCACACTGCCAAAGGCTGCCGTGAAATGTTGATGGAAAGATTTTAAAGGAGCAAACTT 1344  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
 DB 1345 AAAGTCTCCTTCTGCTCGAAGAAAGCTCCAAATGAACAGATGGGGTGTGTGCCACCC 1404  
 QY 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro 480  
 DB 1405 CGTGAAGGACAGGAGCATGACACCACTCCGTGAGGTTCACAGAGCCACAGAGGTCTCT 1464  
 QY 481 GlyGlyProMetGlyArgMetGlyArgGlyGlyLysAspArgGlyGlyLysPheProArg 500  
 DB 1465 GGGGGACCATGATGATGAGAGGCGGTGAGAGGATGAGAGAGGCTTCCCTCCAAAG 1524  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnIleArgAlaGly 520  
 DB 1525 GGAACCCGGGGTTCCTCGAGAGAACCCCTTGAAGAGAGAAAGTCCAGACAGAGCTGGA 1584  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThrGlyCys 540  
 DB 1585 GACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAGAACTTGCCCTGAGAGAACAGAGTGC 1644

QY 541 AaenGlnCybelyaAlaProlyseProGlnGlyPheLeuProProPheProProGly 560  
 DB 1645 AACCAATGTAAGGCCCCAAGGCTTCTCCCGCACCTTCCGCCCCGGGT 1704  
 QY 561 G1yAspArgG1yArgG1yG1yProG1yG1yMetArgG1yG1yArgG1yG1yLeuMetAsp 580  
 DB 1705 GGTGATCGTGAGAGAGGTGCGCTGGTGCAATCGGGAGAAAGAGGTGCGCTCATGAT 1764  
 QY 581 ArgG1yG1yProG1yG1yMetPheArgG1yG1yArgG1yG1yAspArgG1yG1yPheArg 600  
 DB 1765 CGTGCGTGCTCCCGGTGAATGTTCAAGAGGTGCGCTGGTGAAACAGAGGTGCTCCGT 1824  
 QY 601 G1yG1yArgG1yMetAspArgG1yG1yPheG1yG1yG1yArgG1yG1yProG1yG1y 620  
 DB 1825 GGTGGCCGGGGCAATGACCAAGGTGCTTGTGTGAGAGAGAGAGGTGCGCTGGGGGG 1884  
 QY 621 ProProG1yProLeuMetG1yG1yMetG1yG1yArgArgG1yG1yArgG1yG1yProG1y 640  
 DB 1885 CCCCCGTGACCTTGTATGCAACAGATGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1944  
 QY 641 LygMetAspArgG1yG1yMetArgG1yG1yArgG1yG1yArgG1yG1yArgG1yG1y 656  
 DB 1945 AAATGATTAAGGCGAGCAGCGTCAGAGAGCGAGAGATCGCCCTAC 1992

RESULT 2  
 ADQ86032  
 ID ADQ86032 standard; cDNA, 2372 BP.  
 AC ADQ86032;  
 XX  
 DT 07-OCT-2004 (first entry)  
 DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.  
 XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;  
 KM cancer; cell proliferative disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004060270-A2.  
 XX  
 PD 22-JUL-2004.  
 XX  
 PF 15-OCT-2003; 2003WO-US029126.  
 XX  
 PR 18-OCT-2002; 2002US-0418988P.  
 XX  
 PA (GENTH ) GEMENTECH INC.  
 XX  
 PA (WUTD/) MU T D.  
 XX  
 PA (ZHOU/) ZHOU Y.  
 XX  
 PI Wu TD, Zhou Y;  
 XX  
 DR MPI; 2004-534300/51.  
 XX  
 PT New nucleic acid molecule and encoded polypeptide, for diagnosing,  
 XX preventing or treating cell proliferative disorders such as cancer.  
 PS  
 XX  
 PS Claim 1: SEQ ID NO 2904; 5504bp; English.  
 XX  
 CC The present invention describes an isolated tumour-associated antigenic  
 CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
 CC sequences (see SEQ ID NO.1 to 4622); (b) the full-length coding region of  
 CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
 CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
 CC (c). Also described: (1) an expression vector comprising the above  
 CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
 CC a process for producing a polypeptide; (4) an isolated polypeptide  
 CC comprising: (a) an amino acid sequence encoded by any of the above  
 CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
 CC length coding region of the above nucleotide sequences; or (c) a sequence  
 CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
 CC comprising the above polypeptide fused to a heterologous polypeptide; (6)

CC an isolated antibody that binds to the above polypeptide; (7) a process  
 CC for producing the antibody; (8) an isolated oligopeptide that binds to  
 CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
 CC binding organic molecule that binds to the above polypeptide; (10) a  
 CC composition of matter comprising the above (chimeric) polypeptide,  
 CC antibody, oligopeptide or TAT binding organic molecule, in combination  
 CC with a carrier; (11) an article of manufacture comprising a container and  
 CC the composition of matter contained within the container; (12) methods of  
 CC inhibiting the growth of a cell that expresses the above protein, where  
 CC the growth of the cell is at least in part dependent upon a growth  
 CC potentiating effect of the above protein; (13) a method of  
 CC therapeutically treating a mammal having a cancerous tumour comprising  
 CC cells that express the above protein; (14) a method of determining the  
 CC presence of a protein in a sample suspected of containing the protein  
 CC described above; (15) methods of diagnosing the presence of a tumour in a  
 CC mammal; (16) a method for treating or preventing a cell proliferative  
 CC disorder associated with increased expression or activity of the above  
 CC protein; and (17) a method of binding an antibody, oligopeptide or  
 CC organic molecule to a cell that expresses the protein described above.  
 CC The TAT sequences have cytostatic activities, and can be used in gene  
 CC therapy. The composition and methods are useful for diagnosing,  
 CC preventing or treating cancer. The composition is also used for preparing  
 CC a medicament for the therapeutic treatment or diagnostic detection of a  
 CC cell proliferative disorder or cancer. The present sequence represents a  
 CC human TAT cDNA sequence from the present invention.

XX  
 SQ Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;  
 XX

Alignment Scores:  
 Pred. No.: 1,69e-116 Length: 2372  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-10-791-017A-2 (1-656) x ADQ86032 (1-2372)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 44 ATGGCGTCCACGGATTACAGTACTTATAGCAAGTCAGCGAGCGGCTACAGTGT 103  
 QY 21 TyrThrAlaGlnProThrGlnGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 40  
 DB 104 TACACCGCCGACGCTCAAGATATGACACACACACCCAGGCAATGAGCAACAAAGC 163  
 QY 41 TyrGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60  
 DB 164 TATGAACTTATGAGACAGCCCACTGATGTCACTATACCCAGGCTCAGACACTGCAAC 223  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrPro 80  
 DB 224 TATGGGCAAGCCGCTATGCACTTCTTATGACAGCTCCCACTGTTATCTACTCCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100  
 DB 284 ACTGCCCCCGAGCAATACAGCCAGCTGTCCAGAGGATATGGCACTGTATGATAC 343  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 DB 344 ACCACTGCTACAGTACACACACACCGGCTCTTATGACCTAGCTGTGATATGGCACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 DB 404 CAGCGCTTATCAGCACTATGAGGAGAGCCAGCAGCAGCACTCAGCACTCAACAAGCGCAG 463  
 QY 141 ArgGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160  
 DB 464 GATGAAACAAAGCCACTGAGTAATCAACCTCAATTAAGCAGAGGGGTTCACACAG 523  
 QY 161 ProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 524 CCAGCTTAGATATATGACAGAGTAACTACATTAATCCAGATTAATCCAGATTAATCCAG 583

QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 584 ATGAGGCACTGACTGACCTCCACTCTACCTACAGACTATCTCTACAGACGG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 Db 644 ACTAGTATATATGAGACAGTACTCTCAGACAGAACCTTATGGGCAACGAGAGACTAT 703  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 Db 704 GGCAGACAGAGTGTGCTATGCTCAACAAAGCAGCTATGGCAGCCCTCCACTAGTTAC 763  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 764 CCACCCCAACTGATCTTACAGCCCAAGCTCCAACTATGACCAACAGAGCAGCAGC 823  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 Db 824 TACGGGCAAGAGATTCAATCCGACAGAACCCCGATGACATGGGTGTTATGGCAG 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 884 GAGCTGAGAGATTTTCCGAGACAGAGAACCCGAGCATGAGTGCCCTGATTAACCGG 943  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 Db 944 GGCAGGGGAAAGAGGGATTTGATCGTAGAGCATGAGAGAGTGGCCGGAGAGGGA 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnArgProGlyGlyProMet 340  
 Db 1004 CGGGGTGGATGGGCAAGCGCTGAGAGCAGAGTGGCTTCAATAAGCTGTGGAGCCCATG 1063  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLeuAspSerAspAsn 360  
 Db 1064 GATGAAAGACACAGATCTTGATCTAGAGCCACCTGTATATCCAGATGAAAGCTCTGACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnArgSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCATTTATGTACAGATTTAAATGACAGTGTGACTTAAATGATCTGGCAGACTTC 1183  
 QY 381 PheArgGlnGlyValValValMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 Db 1184 TTTAAGCAGATGTGGGCTTTGTTAAGATGACAAAGAACTGGGCAACCATGATCCATC 1243  
 QY 401 TyrLeuAspLysGlyLeuThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLeuAspPro 420  
 Db 1244 TACCTGACAAAGAAAGAAAGAAAGCCAAAGCGATGCCACAGTCTTATGAAGACCA 1303  
 QY 421 ProThrAlaLysAlaAlaValAlaGlyLysPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 Db 1304 CCCACTCCCAAGGCTGCGCTGGAATGCTTGAATGGAAAGATTTTCAAGGAGCAAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
 Db 1364 AAGGTCTCCCTTGTGCGAAGAGCCTCCATGAAGATTTGGGGGTGTGTGCCACCC 1423  
 QY 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480  
 Db 1424 CGTGAAGGAGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCCCCAGAGGCTCT 1483  
 QY 481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500  
 Db 1484 GGGGGAACCAATGGTCCCATGGAGGCGCTGAGAGAAATGAGAGGCTTCTCCCAAGA 1543  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520  
 Db 1544 GGAACCCCGGGGTTCCGAGGGGAACCCCTCTGAGAGAGAAACGTCCAGACCCAGAGCTGA 1603  
 QY 521 AspTyrGlnCysProAspProGlyCysGlyAsnGlnAsnPheAlaTyrPheThrGlnCys 540  
 Db 1604 GACTGCAAGTGTCCCAATCCGGGTTGTGAAACCAACATTCCTGCTGAGAAACAGAGTGC 1663

QY 541 AsnGlnCysLeuAlaProLysProGluGlyPheLeuProProPheProProGly 560  
 Db 1664 AACCATGTAAAGGCCCAAGCCTGAAGGCTTCTCCGCCACCTTCCGGCCCCGGGT 1723  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 Db 1724 GGTGATCGTGGCAGAGGTGGCCCTGCTGACATGCGGGGAGAAAGAGTGGCTTCATGAT 1783  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyArgGlyGlyAspArgGlyGlyPheArg 600  
 Db 1784 CGTGTGTCTCCGATGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTTCGT 1843  
 QY 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 620  
 Db 1844 GGTGGCCGGGCGATGAGCCGAGTGGCTTGTGTGAGAGAAAGAGTGGCTTGGGGGG 1903  
 QY 621 ProProGlyProLeuMetGluGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640  
 Db 1904 CCCCTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGTGGAGACCTGGA 1963  
 QY 641 LysMetAspLysGlyGlyLeuHisArgGlnGlyArgArgAspArgProTyr 656  
 Db 1964 AAATGATTAAGCGAGACCTGTGAGAGCGCAGAGATCGCCCTTAC 2011  
 RESULT 3  
 ID ABK84628 standard; cDNA, 2390 BP.  
 AC ABK84628;  
 XX  
 DT 14-AUG-2002 (first entry)  
 XX  
 DE Human cDNA differentially expressed in granulocytic cells #1199.  
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; chondrosis;  
 KW cardiac respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 OS Homo sapiens.  
 XX  
 PN MO200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001, 2001MO-US030821.  
 XX  
 PR 03-OCT-2000, 2000US-0237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Baezzer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 XX  
 DR WPI, 2002-435328/46.  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.  
 PS Claim 1, SEQ ID NO 1199, 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where  
 CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) GA by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a





QY 241 ProProGlnThrGlySerGlyrSerGlnAlaProSerGlnTrpSerGlnGlnSerSerSer 260  
 DB 764 CCAACCCAACTGATCTCAAGCAAGCTCCAAAGTAATATGACCAAGAGAGAGAGC 823  
 QY 261 TGTGlyGlnGlnSerSerPheArgGlnAAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 824 TAGGGGAGAGAGATTCAATCCAGACAGACCAACCCAGTAGCATGGGTGTTTATGGGAG 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerSerGlyProAspAsnArg 300  
 DB 884 GAGTCTGGAGGATTTTCCGAGCCAGAGAGAACCCGAGCATGATGGCCCTGATTAACGG 943  
 QY 301 GLYArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly 320  
 DB 944 GGCAGGGGAGAGAGGGGATTGATGCTGAGGAGCATGAGAGAGTGGGCGGGAGAGAGA 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 DB 1004 CGGGGTGGATGGGACGCGCTGAGAGAGAGAGTGGCTTCAATAGACCTGGTGGACCAATG 1063  
 QY 341 AspGlyGlyProAspLysAspLysGlyProProValAspProAspGlyLysAspSerAspAsn 360  
 DB 1064 GATGAAGAGACAGATCTTGATCTAGGCGCTCTGTATGATCCAGATGAAGACTGACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlyLysAsnAspSerValThrLysAspLysAlaAspPhe 380  
 DB 1124 AGTGCAATTTATGACAGAGATTAAATGACAGTGTGCTAGATGATCTGGGAGACTTC 1183  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1184 TTTAAGCAGTGGGGTGTGTTAAGATGAACAAGAACTGGGCAACCAATGATCCATC 1243  
 QY 401 TyrLysAspLysGlyGlyLysProLysGlyLysAspAlaThrValSerTyrGlyLysAspPro 420  
 DB 1244 TACCTGGACAAAGAAACAGAAAGCCCAAGGCGCATCCAGTGTCTATGAAGACCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGlnTrpPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1304 CCACTGCGCAAGGCTGCGTGGAAATGATGATGGGAAAGATTTTCAAGGAGAGAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLysProPro 460  
 DB 1364 AAGGTCTCTCTTCTCGAAGAGAGAGCTCCATGAACATGATCGGGGTGATCTCCACCC 1423  
 QY 461 ArgGlyLysArgGlyMetProProProLysArgGlyGlyProGlyGlyProGlyPro 480  
 DB 1424 CGTAGGGGAGAGAGATGACCAACCACTCGTGGAGGTCCAGAGAGGCCAGAGAGTCT 1483  
 QY 481 GLYGlyProMetGlyArgMetGlyGlyArgGlyGlyLysAspArgGlyGlyPheProProArg 500  
 DB 1484 GGGGAGACCCATGGGTGCGCATGGAGGCGCGTGGAGAGATGAGAGAGGCTCTCCCAAGA 1543  
 QY 501 GlyProArgGlySerArgGlyLysProSerGlyGlyGlyLysValGlnHisArgAlaGly 520  
 DB 1544 GAGACCCCGGGGTTCCCGAGGAGAACCTCTCGAGAGAGAAAGCTCCAGCACCAAGCTGGA 1603  
 QY 521 AspTrpGlnCysAspProAsnProGlyGlyGlyLysGlnAsnGlnAsnAlaTrpArgTrpGly 540  
 DB 1604 GACTGGCAGATGCTCCATCCGGGTGGGAAACAGAACTTCCGCTGGAACAAGAGTGC 1663  
 QY 541 AsnGlyCysLysAlaProLysProGlyGlyPheLeuProProProPheProProProGly 560  
 DB 1664 AACCAAGTAAAGGCCCAAGAGCCGAAAGGCTTCTCCCGCACCTTCCGCGCCCGGAT 1723  
 QY 561 GLYAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyLysArgGlyGlyLysMetAsp 580  
 DB 1724 GGTGATGTGGCAGAGGTGCGCTGTGGCATGGGGGAGAGAGAGTGGCTCATAGAT 1783  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyLysArgGlyGlyLysAspArgGlyGlyPheArg 600  
 DB 1784 CGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCGGTGTGGAAGCAAGAGTGGCTTCGT 1843  
 QY 601 GLYGlyArgGlyMetAspArgGlyGlyPheGlyGlyLysArgArgGlyGlyProGlyGly 620

DB 1844 GTGGCCGGGGCATGAGCCAGAGTGGCTTGTGGAGAGAAAGAGAGTGGCCCTGGGGGG 1903  
 QY 621 ProProGlyProLysMetGlnGlnMetGlyGlyLysArgArgGlyGlyArgGlyGlyProGly 640  
 DB 1904 CCCCCTGAGACTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlyGlyLysHisArgGlnGlyLysArgArgAspArgProTyr 656  
 DB 1964 AAATGGATTAAGCGCAGACCGCTCAGAGAGAGAGAGATGGCCCTTAC 2011  
 RESULT 5  
 ID ABN97274 standard; DNA; 2390 BP.  
 AC ABN97274;  
 DT 13-AUG-2002 (first entry)  
 DE Gene #3772 used to diagnose liver cancer.  
 XX Gene; liver cancer; de; hepatocellular carcinoma; hepatotropic;  
 KM metastatic liver tumour; cytostatic; expression profile; disease state;  
 KM disease progression; drug toxicity; drug efficacy; drug metabolism.  
 XX Homo sapiens.  
 OS  
 PN WO200229103-A2.  
 PD 11-APR-2002.  
 PF 02-OCT-2001; 2001MO-US030589.  
 XX  
 PR 02-OCT-2000; 2000US-0237054P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 XX WPI; 2002-426119/45.  
 DR  
 XX  
 PT Diagnosing and detecting the progression of liver cancer, hepatocellular  
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
 PT level of expression of two or more genes in a liver tissue sample.  
 XX  
 PS Claim 1; SEQ ID NO 3772; 298bp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 CC  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1,7e-116 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-791-017A-2 (1-656) X ABN97274 (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrTyrSerAla 20  
 DB 44 ATGGCGGCCAGCGATTACAGTACCTATAGCCAGCTGACGCGAGCGGGCTCAAGGCT 103  
 QY 21 TyrThrAlaGlnProThrGlnGlnTyrTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 40  
 DB 104 TACACGCCGCCAGCCCACTCAAGATATGACAGACCCAGGCAATATGGCCAAACAGC 163  
 QY 41 TyrGlnTyrThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 DB 164 TATGGAACCTATGACAGCCCACTGATGTCACTATACCAAGGCTCAGACCACTGCAACC 223  
 QY 61 TyrGlnGlnThrAlaTyrAlaThrSerTyrGlnProThrProTyrTyrThrThrPro 80  
 DB 224 TATGGGAGACCGCTATGCAACTCTTATGAGACAGGCTCCCACTGGTTATACCTCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlnTyrGlnTyrAlaTyrAspThr 100  
 DB 284 ACTGCCCCCAAGGATACAGCCAGCTGTCCAGGGGATATGGCACTGGTCTTATGATACC 343  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlnTyr 120  
 DB 344 ACCACTGCTACAGTCAACACCAACCCAGGCTCTATGACCTCAGTCTGATATGCACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 DB 404 CAGCTGCTTATCCAGCTTATGGGACAGCAGCAGCAGCAGCAGCTGACCTTCAAGACCCGAG 463  
 QY 141 AspGlyAsnLysProThrGlnTyrThrSerGlnProGlnSerThrGlnTyrTyrAsnGln 160  
 DB 464 GATGGAAACAAGCCCACTGAGACCTATCACTCACTCACTGACAGGGGTTTACAAACAG 523  
 QY 161 ProSerLeuGlnTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 524 CCCAGCTTATGAGATGAGACAGAGTAACTCAAGTATCCCAAGGTAACCTGGGAGCTACCCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 584 ATGAGCGCACTGACCTGACCTTACCTTCACTCACTCACTCACTTCTTCAACAGCCG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlnGlnProSerSerTyr 220  
 DB 644 ACTAGTATGATGACAGACAGTACTCTCAGAGAAACCTATATGGGACCAAGAGACGATAT 703  
 QY 221 GlnGlnGlnSerTyrGlnGlnGlnSerTyrGlnGlnGlnProProThrSerTyr 240  
 DB 704 GGAACAGAGAGTATGATGATCAACAAAGCAAGCTATGGGACAGCTCCCACTAGTTAC 763  
 QY 241 ProProGlnThrGlnSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260  
 DB 764 CCAACCCCAACTGATCTTACAGCCAGCTCAAGTCAATATATGCAACAGAGAGAGCAGC 823  
 QY 261 TyrGlnGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlnGln 280  
 DB 824 TAGGGGACAGAGATTTCATCCAGACAGACCAACCCAGTATGATGGTGTATTATGGGACG 883  
 QY 281 GlnSerGlnGlnPheSerGlnProGlnGlnAlaAsnArgSerMetSerGlnProAspAsnArg 300  
 DB 884 GAGCTGTGAGAGATTTTCCGAGCCAGAGAGAAACCGGAGCATATGATGCTCCCTGATACCGG 943  
 QY 301 GlnTyrGlnTyrArgGlnPheAspArgGlnGlnMetSerArgGlnTyrArgGlnGlnTyr 320  
 DB 944 GGCAGGGGAAAGAGGGGATTTGATCTGAGAGGCAATGACAGAGTGGGCGGAGAGAGGA 1003  
 QY 321 ArgGlnTyrMetGlnSerAlaGlnGlnTyrGlnTyrPheAsnLysProGlnTyrProMet 340  
 DB 1004 CGCGGTGGAATGGGCAAGCGCTGAGAGACAGAGTGGCTTCAATAAGCTGTGATGCCATG 1063  
 QY 341 AspGlnGlnTyrProAspLeuAspLeuGlnTyrProProValAspProAspGlnAspSerAspAsn 360  
 DB 1064 GATGAAGGACCAAGATCTTGAATCTAGGCCCTCTCTATAGATCCAGATGAAGACTCTGACAA 1123

QY 361 SerAlaIleTyrValGlnGlnGlnLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1124 AGTGCATTTTATGATACAGAGATTTAAATGACAGTATCACTTATGATATCTGGACAGCTTC 1183  
 QY 381 PheLeuGlnCysGlnValValLysMetAsnLysArgThrGlnGlnProMetIleHisIle 400  
 DB 1184 TTTAAACAGATGGGGTGTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCACATC 1243  
 QY 401 TyrLeuAspLysGlnThrGlnLysProLysGlnLysAspAlaThrValSerTyrGlnAspPro 420  
 DB 1244 TACCTGACACAGAAACAAGAAAGCCCAAGGGGATGCCACAGTGTCTTATGAAAGACCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGluTyrPheAspGlnLysAspPheGlnGlnSerLysLeu 440  
 DB 1304 CCCACTGCCAAGGCTCTCCGTGAAATGTTTATGAGAAAGATTTTCAAGGAGACAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlnTyrLeuProPro 460  
 DB 1364 AAAGTCTCCCTTGTCTCGAAGAAAGCTCCCAATGAACAGTATGGGGGTGTGCCACCC 1423  
 QY 461 ArgGlnGlnTyrArgGlnMetProProProLysArgGlnTyrProGlnTyrProGlnTyrPro 480  
 DB 1424 CGTGAAGGACAGGACATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAGGTCTT 1483  
 QY 481 GlnTyrProMetGlnTyrMetGlnTyrArgGlnTyrAspArgGlnTyrPheProProArg 500  
 DB 1484 GGGGAGACCAATGGGTGTGATGGAGGCGCTGAGAGAGATAGAGAGCTTCCCTCCAAAG 1543  
 QY 501 GlnProArgLysArgLysArgLysAsnProSerGlnTyrGlnLysValGlnHisArgAlaGln 520  
 DB 1544 GAGCCCCGGGGTTCGGAGGGAACCCCTCTGAGAGAGGAACGCTCCAGCACCGAGCTGGA 1603  
 QY 521 AspTyrGlnCysPheAsnProGlnTyrCysGlnLysGlnAsnPheAlaTTPATGTTGluCys 540  
 DB 1604 GACTGGCAGATGTCCTCAATCCGGGTGTGGAACCAACACTTCCCTGAGAAACAGAGTGC 1663  
 QY 541 AsnGlnCysLeuAlaProLysProGlnGlnTyrPheLeuProProProPheProProProGln 560  
 DB 1664 AACCACTGTAAAGCCCAACAGCTGAAAGCTTCTCCGCCACCTTCCGCCCGGGGT 1723  
 QY 561 GlnAspArgGlnTyrArgGlnTyrProGlnTyrMetArgGlnTyrArgGlnTyrLeuMetAsp 580  
 DB 1724 GGTGATCTGTGGCAGAGGTGGCCCTGTGTGACATGCGGGAGAAAGTGGCTCATGAT 1783  
 QY 581 ArgGlnTyrProGlnTyrMetPheArgGlnTyrArgGlnTyrGlnAspArgGlnTyrPheArg 600  
 DB 1784 CGTGTGTGTCCTCCGTGAATGTTCAAGAGGTGGCCGTGTGGAACAGAGGTGGCTTCCGT 1843  
 QY 601 GlnTyrArgGlnMetAspArgGlnTyrPheGlnTyrGlnTyrArgArgGlnTyrProGlnTyr 620  
 DB 1844 GGTGGCCGGGCAATGAAACGAGGTGGCTTGTGTGAGAGAAAGAGAGTGGCCCTGGGGGG 1903  
 QY 621 ProProGlnTyrProLeuMetGlnGlnMetGlnTyrArgArgGlnTyrArgGlnTyrProGln 640  
 DB 1904 CCCCTGGAACCTTATATGGAACAGATGGAGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlnTyrHisArgGlnGlnLysArgArgAspArgProTyr 656  
 DB 1964 AAATATGATTAAGCGAGACCGTCAAGAGCGGAGAGATCGGCCCTTAC 2011

RESULT 6  
 ADS16277  
 ID ADS16277 standard; cDNA, 2390 BP.  
 XX ADS16277;  
 AC  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human cDNA encoding the Ewing sarcoma protein Segid 1.  
 KW human; ss; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
 hypogonadism; androgen-resistance syndrome; testicular feminization.



XX	OS	Hom sapiens.	
XX	Key	Location/Qualifiers	
XX	FT	44..2014	
XX	FT	/tag= a	
XX	FT	/product= "Ewing sarcoma protein"	
XX	FT		
XX	FN	EP1455190-A1.	
XX	PD	08-SEP-2004.	
XX	PF	16-FEB-2004; 2004EP-00003422.	
XX	PR	04-MAR-2003; 2003DE-01009280.	
XX	PR	25-APR-2003; 2003US-0465692P.	
XX	PA	(SCHD ) SCHERING AG.	
XX	PI	Obendorf M, Wolf S;	
XX	DR	WPI; 2004-627861/61.	
XX	DR	P-PSDB; ADS16278.	
XX	PT	Determining the hormonal effects of substances, used to identify	
XX	PT	pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,	
XX	PT	from modulating interaction between nuclear receptors and Ewing sarcoma	
XX	PT	protein.	
XX	PS	Claim 7, SEQ ID NO 1; 30pp; German.	
XX	CC	This invention relates to a novel modulators that alter the interaction	
XX	CC	between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well	
XX	CC	as the screening method thereof. Specifically, it refers to determining	
XX	CC	and identifying a hormonal effect brought about by test compounds that	
XX	CC	modulate either the binding of EMS to the nuclear receptor or the ligand	
XX	CC	induced activity of this receptor. The present invention describes the	
XX	CC	nuclear receptors as including oestrogen, progesterone, thyroid hormone,	
XX	CC	Vitamin D, and retinoic acid receptors, most preferably they are androgen	
XX	CC	receptors. Accordingly, these modulators may be used in the development	
XX	CC	of pharmaceutical compositions that can diagnose and be used to treat	
XX	CC	diseases associated with receptor dysfunction such as prostatic cancer,	
XX	CC	alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.	
XX	CC	testicular feminisation. This method provides reliable, sensitive,	
XX	CC	simple, inexpensive and rapid assessment of the hormonal effects of these	
XX	CC	test compounds. This polynucleotide sequence is the cDNA encoding the	
XX	CC	human Ewing sarcoma protein of the invention.	
XX	XX	Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;	
XX	XX		
XX	XX	Alignment Scores:	
XX	XX	Pred. No.: 1.7e-116 Length: 2390	
XX	XX	Score: 3633.00 Matches: 656	
XX	XX	Percent Similarity: 100.00% Conservative: 0	
XX	XX	Best Local Similarity: 100.00% Mismatches: 0	
XX	XX	Query Match: 100.00% Indels: 0	
XX	XX	DB: 13 Gaps: 0	
XX	XX	US-10-791-017A-2 (1-656) x ADS16277 (1-2390)	
XX	QY	1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20	
XX	Db	44 ATGGCGTCCACGGATTACAGTACTTAACTTAAAGCAAGTCGACGGCGCAAGGCTCAAGTCT 103	
XX	QY	21 TyrThrAlaGlnProThrGlnGlnGlyTyrAlaGlnThrThGlnAlaTyrGlyGlnGlnSer 40	
XX	Db	104 TACACCGCCCAAGCCCACTCAAGATATGCAAGACCAACCAAGCATATGGGCAACAAGC 163	
XX	QY	41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThGlnAlaGlnThrThrAlaThr 60	
XX	Db	164 TATGAGACCTATGAGACGCCCACTATATGTACGTTAAACCAAGGCTCAAGCACTGCAACC 223	
XX	QY	61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80	

Db	224	TATGGGACGACCGCTATGCACTTCTTATGGACAGCCCTCCACCTGGTTATCTCTCA	283
OY	81	ThralaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyValaTyrAspThr	100
Db	284	ACTGCCCCCGACGAGATACACCGACGCTGTCCAGGGGATATGGCACTGGTGTATATGATCC	343
OY	101	ThrThralaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
Db	344	ACCACTGGCTACAGTACCAACCAACCGAGCTCTATGACGCTCAATCTGGCAATGGCACT	403
OY	121	GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln	140
Db	404	CAGCTGTCTTATCCAGCCTATGGGACGACGACGACGACCTGCACTTACAAGACCGCAG	463
OY	141	AspGlyAsnLysProThrGluThrSerGlnProGlnInserSerThrGlyGlyTyrAsnGln	160
Db	464	GATGGAAACAAAGCCCACTGACGACTAGTCAACTCAATCTAGACAGGGGGTTACAAACCG	523
OY	161	ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	524	CCACAGCTTGGATATNGACAGAGTACTACAGTTATCCCAAGTACTCTGGAGCTACCC	583
OY	181	MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200
Db	584	ATGCAGCAGTCACTGCACCTCCATCTCAACCTCCTACACAGTATTCCTTACACAGCG	643
OY	201	ThrSerTyrAspGlnInserSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220
Db	644	ACTAGTTATGATCAGACAGAGTACTCTCAGACAAACACTTATGGGCAACGACAGAGCTAT	703
OY	221	GlyGlnGlnSerSerTyrGlyGlnGlnInserSerTyrGlyGlnGlnProProThrSerTyr	240
Db	704	GGAACGACGAGTAGTAGTACGTCAACAAAGCACTATGGGACAGACCTCCCACTAGTTAC	763
OY	241	ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnInserSerSer	260
Db	764	CCACCCCAACTGGATCCTACAGCCAACTCCAAAGTCAATATGACCAACAGACGACGAGC	823
OY	261	TyrGlyGlnGlnInserSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln	280
Db	824	TACGGGACGACAGTCTATCTCCGACAGGACCAACCCCACTATCAGTAGGATGTTATATGGGACG	883
OY	281	GluSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerMetSerGlyProAspAsnArg	300
Db	884	GAGTCTGAGGATTTTCCGACCCAGAGAGAAACCGAGCATGATGGCCCTGATTAACCGG	943
OY	301	GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly	320
Db	944	GGCAGGGGAAAGGGGGGATTTGATCTGTGAGGATGACAGAGTGCGGGGGAGGAGAGA	1003
OY	321	ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet	340
Db	1004	CGCGGTGAATAGGGCAGCGCTGGAGAGGAGTGGCTTCATTAAGCCCTGGTGGACCCCATG	1063
OY	341	AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn	360
Db	1064	GATGAAGAACCAATCTTGATCTTAGGCCCACTGATGATCCAGATGAAAGACTCTGACAAAC	1123
OY	361	SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe	380
Db	1124	AGTCAATTTATGATGCAAGGATTTAAATGACAGCTGACTCAAGTATGATCTGGACAGACTTC	1183
OY	381	PheLysGlnCysGlyValValLysMetCAsnLysArgThrGlyGlnProMetIleHisIle	400
Db	1184	TTTAAAGCAGCTGGGGTGTATTAATATACAACAAGAACTGGGCAACCATGATCCACATC	1243
OY	401	TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro	420
Db	1244	TACCTGGACAAAGGAAACGAGAAACCCAAAGGCATGACACAGTGTCTTATGAAGAACCA	1303
OY	421	ProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerTyrLeu	440

Db 1304 CCCACTGCAAGCTCCGCGATGTTGATGGAAGATTTTCAGAGGACCAACTT 1363  
 Qy 441 LyeValSerLeuAlaArgIlyIyProProMetAsnSerMetArgIlyIyLeuProPro 460  
 Db 1364 AAAGTCTCCCTTGTCCGAGAAAGCTTCATGAAGATGCGGGTGTGTCTGCCACCC 1423  
 Qy 461 ArgIlyIyArgIlyMetProProProLeuArgIlyIyProIyIyProIyIyPro 480  
 Db 1424 CGTGAGGCGAGAGGACGACCACTCGTGGAGAGTCCAGAGGCCAGAGAGTCCCT 1483  
 Qy 481 G1yIyProMetG1yArgMetG1yIyArgIlyIyAspArgIlyIyPheProProArg 500  
 Db 1484 GGGGGAACCCATGGGTCCGATGGAGGCGCTGAGAGATGAGAGGCTTCCCTCCAGA 1543  
 Qy 501 G1yProArgIlySerArgIlyAsnProSerG1yIyIyIyIyIyIyIyIyIyIyIy 520  
 Db 1544 GAGACCCGGGGTTCCTCCAGGAGAACCTCTGGAGAGAGAAACGTCAGACCGAGCTGGA 1603  
 Qy 521 AspTrpGlnCysProAsnProG1yCysG1yAsnG1yAsnG1yAsnG1yAsnG1y 540  
 Db 1604 GACTGGCAGTGTCCCATTCGGGTGTGGAAACCAACTTCGCTGGAAGACAGAGTCC 1663  
 Qy 541 AsnGlnCysIy 560  
 Db 1664 AACCAAGTAAAGGCCCAAGCCGAAAGGCTTCCCTCCGACACCTTCCGCGCCGCGGT 1723  
 Qy 561 G1yAspArgIlyIyArgIlyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 580  
 Db 1724 GGTATGTGTGCAAGAGTGTGGCTGTGGCAATGGGAGAGAAAGTGTGCTCATGAT 1783  
 Qy 581 ArgIlyIyIyProG1yIyIyMetPheArgIlyIyIyArgIlyIyIyAspArgIlyIyPheArg 600  
 Db 1784 CGTGTGTGTCCCGGTGGAATCTTCAAGGTGTGCGCTGTGGAGACAGAGTGTCTCCCT 1843  
 Qy 601 G1yIyIyArgIlyIyMetAspArgIlyIyIyPheG1yIyIyIyIyIyIyIyIyIy 620  
 Db 1844 GGTGGCGGGGATGAGACCGAGGTGTGTGTGAGAGAAAGAGTGTGCGCTGCGGGG 1903  
 Qy 621 ProProG1yProLeuMetG1yIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 640  
 Db 1904 CCCCCTGACCTTGTGATGAGAACAGATGGAGAGAAAGAGAGACGTGAGACCTGGA 1963  
 Qy 641 IyMetAspIyIyG1yIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 656  
 Db 1964 AAAATGATTAAGCGCAGACCGTCAAGAGCGCAGAGATGCGCCCTAC 2011  
 RESULT 7  
 ACN40903 ID ACN40903 standard; cDNA, 2390 BP.  
 ACN40903;  
 18-NOV-2004 (first entry)  
 Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.  
 Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 tumour; diagnosis; cell proliferative disorder; breast cancer;  
 colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 central nervous system cancer; bladder cancer; pancreatic cancer;  
 cervical cancer; melanoma; leukaemia; hybridisation probe;  
 chromosome identification; chromosome mapping; gene mapping;  
 gene therapy; cytotoxic; gene; ss.  
 Homo sapiens.  
 OS  
 PN WO2004030615-A2.  
 PD 15-APR-2004.  
 XX 29-SEP-2003; 2003WO-US028547.  
 PF 02-OCT-2002; 2002US-0414971P.  
 XX

XX (GETH) GENENTECH INC.  
 PA Wu TD, Zhang Z, Zhou Y;  
 PI MPI; 2004-347921/32.  
 DR P-PSDB; ABM82330.  
 XX  
 PT New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 PS Claim 1; SEQ ID NO 5989; 7273bp; English.  
 XX  
 CC The invention relates to human tumor-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT nucleic acid of the invention  
 XX  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1,7e-116 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 13 Gaps: 0  
 US-10-791-017A-2 (1-656) x ACN40903 (1-2390)  
 Qy 1 MetAlaSerThrAspIyIySerThrIyIySerGlnAlaAlaIyIyIyIyIyIyIyIyIyIy 20  
 Db 44 ATGGCGTCCACGAGATTACAGTACTTATGACCAAGTGCAGCGCAGGCTACAGTGT 103  
 Qy 21 TyrThrAlaGlnProThrGlnIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 40  
 Db 104 TACACCGCCGACGACCTCAAGATATGACACACCAACGACGACATATGGCAACAGC 163  
 Qy 41 TyrG1yIy 60  
 Db 164 TATGAAACCTATGAGACGCCACTGATGATGATATACAGGCTCAGACACTGCAACC 223  
 Qy 61 TyrG1yIy 80  
 Db 224 TATGGCGACGCCCTATGCACTTCTTATGACAGCTCCACTGGTTATATCACTCCA 283  
 Qy 81 ThrAlaProGlnAlaIyIySerGlnProValGlnIyIyIyIyIyIyIyIyIyIyIyIy 100  
 Db 284 ACTGCCCCGAGGATACAGCGACCTGTCCAGGGGATATGGCACTGTATATGATACC 343  
 Qy 101 ThrThrAlaThrValThrThrThrGlnAlaIyIyIyIyIyIyIyIyIyIyIyIyIy 120  
 Db 344 ACCACTGTACATGCCACACCGAGGCTCTTATGCACTCAAGTGTGATATGGCACT 403  
 Qy 121 GlnProAlaIyIyProAlaIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 140



CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. A5662214-A5662388  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins  
XX  
SQ Sequence 2176 BP, 413 A, 617 C, 563 G, 583 T; 0 U; 0 Other;

**SQ Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;**

**Alignment Scores:**

Pred. No.:	1,440-115	Length:	217
Score:	360.5	Matches:	655
Percent Similarity:	99.85	Conservative:	0
Best Local Similarity:	99.85	Mismatches:	1
Query Match:	99.23	Indels:	1
DB:	6	Gaps:	0

US-10-791-017A-2 (1-656) x AAS62262 (1-2176)

Qy	1	MetLAsERTAApPYSeSThTyrSergInAlaAlaAGInGInG1TYSeSLa	20
Db	2160	ATGGCGTCCAGGATTACGTAACCTATAGCCAGGCTGACGACAGGCTACAGTCT	2100
Qy	21	TyrThraLagInProthGInG1TYraLagInThrthGInAlaTYrG1yGInInSer	40
Db	2100	TACACGGCCAGCCACTCAAGATATGCAAGACCACCAGGCTATATGGGCAAAAGC	2041
Qy	41	TyrG1yThrTYrG1yGInProThraAspValSerTYrThrGInAlaGInThThraAlaThr	60
Db	2040	TATGGAACTTATGGACAGCCCACTGATGTCACTATACCCAGGCTCAACCACTGCAAC	1983
Qy	61	TYrG1yGInThraAlaTYraAlaThSerTYrG1yGInProThG1TYrThThrPro	80
Db	1980	TATGGGCAAGCCGCTATGCACTTCTATAGCAGGCTCCCACTGGTATATCTACTCA	1922
Qy	81	ThraLaProGInAlaTYrSerGInProValGInG1TYrG1yThrG1yAlaTYraAspThr	100
Db	1920	ACTGCCCCCAAGCAATACAGCCAGGCTGTCCAGGGGTATGGCACTATGGCTTATATAC	1864
Qy	101	ThrThraAlaThraValThrThrThrGInAlaSerTYraAlaLagInSerAlaTYrG1yThr	120
Db	1860	ACCACTGTATACATCACCAACCAACCAGGCTCTCATATGACGCTCAATCTGCATATATGGCACT	1801
Qy	121	GInProAlaTYrProAlaTYrG1yGInGInProAlaAlaThraLaProThraPProGIn	140
Db	1800	CAGCTGCTTATCCAGGCTATGGGCAAGCAGCAGCACTGCACCTCAACAGCCGCA-	1744
Qy	141	AspG1yAsnLysProthG1uThSerGInProGInSerSerThG1yG1TYraGIn	160
Db	1741	GATGGAAACAAAGCCCACTGAGACTATGTAACCTCAATCTAAGACAGGGGGTTACAAACAG	1682
Qy	161	ProSerLeuG1yTYrG1yGInSerAsnTYrSerTYrProGInValProG1ySerTYrPro	180
Db	1681	CCGAGCCCTAGATATGACAGAGTAACATACGTTATCCCAAGTACCTGGAGGCTACACCC	1622
Qy	181	MetGInProValThraLaProProSerTYrProProThSerTYrSerSerThGInPro	200
Db	1621	ATGCAGCCAGTCACTGCACCTTCATCTACCTCTTACAGCTATTCCTTACACACCCG	1566
Qy	201	ThrSerTYraAspGInSerSerTYrSergInGInAsnThTYrG1yGInProSerSerTYr	220
Db	1561	ACTAGTTATATGACAGGACTTACTCTCAGCAGAACTATATGGGCAACCGAGGAGCTAT	1502
Qy	221	G1yGInGInSerSerTYrG1yGInGInSerSerTYrG1yGInGInProProThSerTYr	240

Db	1501	GGACAGCAGAGTACTACTATGCTCAACAAAGCAGCTATGGGCGACGACCTCCACTAGTTAC	1444
Oy	241	ProProGlnthnrglyyseryserglnalaproserglnthrsrglnglnserser	260
Db	1441	CCACCCCAATGTGATCTCTACAGCCAAAGCTCAAGTCAATATAGCAACAGACGACGC	1385
Oy	261	TyrglynglnserserPhearglnasphiasProsersermetglyValtyrtygln	280
Db	1381	TACGGGACGACGAAATTCTATCCGACAGAACACCCCACTACATGGGTGTTATGGGCAG	1322
Oy	281	GluserglyglyPheSerglyProglyValuanarysermetSerglyProasphanarg	300
Db	1321	GAGTCTGGAGGATTTTCCGAGCCAGAGAGAACCGAGACATGAGTGCGCTTGATTAACGG	1265
Oy	301	GlyargrglyargglyglyPheasphargglyglymetserargglyglyValargglygly	320
Db	1261	GGCAGGGGAAAGGGGGGATTTGATCTGGAGAGCATGACAAAGTGCGCGGGGAGAGAGA	1205
Oy	321	ArgglyglymetGlyseralaglValuanrgglyglyPheanlnyProglyglyPromet	340
Db	1201	CGCGGTGGAATGGGCGAGCGCTGGAGAGCGAGGTGGCTTCAATTAAGCTGTGTGACCCATG	1144
Oy	341	AspgluglyProaspleuaspleuglyProProValaspproaspgluaspsersaphan	360
Db	1141	GATGAGAGGACCAAGATCTTGATCTAGCGCCACCTGTAGATCCAGATGAGAGACTGTCAAC	1085
Oy	361	SeralalertyValglnglyVleuasnaspservalThrleuasnaspleuanlAasPhe	380
Db	1081	AGTCCAAATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATGTGGCAACTTC	1022
Oy	381	PheylasglnCysglyValVallyMetAsnlyasargThrarglyValnPrometllethstlle	400
Db	1021	TTTAAGCAGTGTTGGGGTGTGTTAAATATGAAACAAAGAACTGGGCAACCCATATGCCACTC	962
Oy	401	TyrluasnaspleugluthnrglyLysasprolysglyAspalathrvalserTyrglyuasppro	420
Db	961	TACTGTGACAAAGAAACAGAAABCCCAAGAGCATGCCAGATGCTCTATGAAACACCA	902
Oy	421	ProThrAlalyValaalavalGluThrPheaspglylyAsasppheglnglyserlyseu	440
Db	901	CCCACTGGCAAGGCTCGCTGGTAATGTGTTATGGGAAAGATTTTCAAGGGGCAAACTT	842
Oy	441	LysValserleuAlaarglyVelyProPrometAsnsermetArgglyVleuProPro	460
Db	841	AAAGTCTCCCTGTCTCGGAAAGACCTTCAATGAAACGTATGGGGGGTGTGCAACC	782
Oy	461	ArggluglyargglymetProProProleuasargglyglyProglyglyProglyglyPro	480
Db	781	CGTGAAGGACAGGACATCCACACACACTCTCGTGGAGGTCCAGAGAGGCCAGAGAGTCTT	722
Oy	481	GlyglyProMetCglyargmetCglyglyValargglyglyAsasargglyglyPheProProArg	500
Db	721	GGGGGACCCCAATGGGTCGATGGAGAGCCGTGAGAGAGATTAAGAGAGGCTTCCCTCCAGA	662
Oy	501	GlyProAsargglyserargglyAsasProserCglyglyValasnvAlGlnHisArgglylagly	520
Db	661	GGACCCCGGGGTTCCCGAGGAACCCCTCTGAGAGAGAAACGTCCACACCGAGACTGGA	602
Oy	521	AspTrpGlnCysProAsnProglyCysglyVasnglnnasPheAlaTrpArgThrGlnCys	540
Db	601	GACTGGCAGTGTCCCAATCCGGGGTGTGTGAAACCAAGAACTTCGCTGAGAAACAGAGTGC	542
Oy	541	AsnglnCyslyValaProlyasProglyuglyPheleuProProProPheProProProgly	560
Db	541	AACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCCGCCGCAACCTTCCGGCCCCCGGGT	482
Oy	561	GlyAsasargglyargglyglyProglyglylymetasargglyglyValargglyglylyleumetAsp	580
Db	481	GGTATCTGTGCAAGGTGGCCCTGTGGGCATGTGGGAGAGAAAGGTGGCTCTCATGTGAT	422
Oy	581	ArgglyglyProglyglylymetPheasargglyglyValargglyglylyAsasargglyglylyPheArg	600
Db	421	CGTGGTGTCCCGGGTGAATGTTCAGAGTGGCCGTGTGTGAAGACAAAGTGGCTTCCGT	362





QY 59 aThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProthrglyTyrThrTh  
 Db 233 AACCTAATGGGAGACCCGCTAATGCAACTTCTTAATGACAGCCCTCCACTGGTATAC  
 QY 79 rProThrAlaProGlnAlaTyrSer-GlnProValGln-GlyTyrGlyThrGlyAlaTyr  
 Db 293 TCCAACTGCCCCCAGGATACAGCCAGCCTGTCAGGGGATATGGCATGGTCTTAT  
 QY 99 AspThrThrThrAlaThrValThr-ThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyr  
 Db 353 GATACCACTGCTACAGTCAACCAACCAAGCCCTCCATGCAAGTCACTGTCATA  
 QY 118 rGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAr  
 Db 413 TGGCACTCAAGCTCTTATCAAGCTATGGGACAGCAGCAAGCACTTCCCTACAG  
 QY 138 gProGlnAsp-GlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyr  
 Db 473 AACGAGATTTGAAACAGCCCACTGAGACTAGTCAACTCAATCTAGCAAGAGGAGT  
 QY 158 YrAnglnProSerLeuGlyTyrGlyGlnSerAenTyrSerTyrProGlnAlaProGlys  
 Db 533 ACAACCAAGCCAGCTGAGATGACAGAGTAACTAAGTATCCCAAGTACTGGGA  
 QY 178 eTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerT  
 Db 593 GCTAACCCATGACAGCCAGTCACTGCACTCCATCCCTCAACAGCTATTCCTCTA  
 QY 198 hrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAenThrTyrGlyGlnPro  
 Db 653 CACAGCCGACTAGTATGATGATGACAGCTACTCTCAGCAAAACCTATGAGGACCGA  
 QY 218 eSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerTyrGlyGlnGlnProProT  
 Db 713 GCAGCTATGACAGCAGAGTATGATGATGATCAAAAGAGTATGGGACAGACTCCCA  
 QY 238 hrSerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGln  
 Db 773 CTAGTATCCCAACCCCAACCTGATCTCAACAGCCCAAGCTCCCAATATGACCAAGA  
 QY 258 eSerSerTyrGlyGlnGlnSerSerPheArgGlnAenPheProSerSerMetGlyValT  
 Db 833 GCAGCACTACGGGACAGAGTTCATTCAGACAGAACCAACCCAGTACATGGGTCTT  
 QY 278 YrGlyGlnGlnSerGlyGlyPheSerGlyProGlyGlnAenArgSerMetSerGlyProA  
 Db 893 ATGGGACGAGAGTCTGAGAGATTTCCGGAACAGAGAGAACCGGACATGATGCGCTG  
 QY 298 ePheAenArgGlyArgGlyArgGlyGlyPheAenPheArgGlyGlyMetSerArgGlyValArg  
 Db 953 ATAAACCGGGGACAGGAGAGAGGAGATTTGATCGTGAAGCATAGCAGATGGCGGG  
 QY 318 YrGlyGlyArgGlyGlyMetGlySerAlaGlyGlyGlyGlyGlyPheAenLysProGlyG  
 Db 1013 GAGGAGGACCGGGATGAGATGGGACGCTGAGAGCAAGGAGGCTTCAATAAGCTGCTG  
 QY 338 1yProMetAspGlnGlyProAspLysLeuAspLeuGlyProProValAspProAspGlyLys  
 Db 1073 GACCCAGTGAAGAGACAGATCTTATCTAGGCCCTCTCTGATGATCCAGATGAACACT  
 QY 358 ePheAenSerAlaLeuTyrValGlnGlyLeuAenAspSerValThrLeuAspAenLeu  
 Db 1133 CTGCAACAGTGCATTAATTAATGACAGATTAATGACAGTGTACTCTGATGATCTGG  
 QY 378 1aAspPhePheLysGlnCysGlyValValLysMetAenLysArgThrGlyGlnProMetI  
 Db 1193 CAGACTCTTTAAGCAGAGTGTGGGTGTTAAGATGAACAAGAGAACTGGGCAACCCATGA  
 QY 398 1eHisLeuTyrLeuAspLysGlyLysThrGlyLysProLysGlyAspAlaThrValSerTyrG  
 Db 1253 TCCGACATCTACTGGACAGAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTTATG

QY 418 1uAspProProThrAlaLysAlaValGluTyrPheAspGlyLysAspPheGlnGlyS  
 Db 1313 AAGACCAACCAACGAGCCAGAGCTGCCGTGAAATGATGATGGGAAAGATTTTCAAGGA  
 QY 438 eLysLeuLysValSerLeuAlaArgLysLysProProMetAenSerMetArgGlyLysL  
 Db 1373 GCAACTTAAGTCTCCCTTGTCTCGAAGAAAGCTCCAAATGAACATATGGGGATGTC  
 QY 458 eUpProAspArgGlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProG  
 Db 1433 TGGCAACCCGAGAGGAGGAGGAGGATGCCAACCAACCTCCGAGAGTCCAGAGAGCCAG  
 QY 478 YrGlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheP  
 Db 1493 GAGGTCTGGGGGACCATGGGTGCAATGGGAGGCGCTGAGAGAGATAGAGAGAGCTTCC  
 QY 498 rOProArgGlyProArgGlySerArgGlyAenProSerGlyGlyGlyAsnValGlnHisA  
 Db 1553 CTCCAAAGAGACCCCGGGGTTCCCGAGAGAACCTCTTGAAGAGAGAACGTCACAGC  
 QY 518 rGAlaGlyAspTyrGlnCysProAenProGlyCysGlyAenGlnAenPheAlaTyrArgT  
 Db 1613 GAGCTGAGACTGGCACTGTCCCAATCCGGGTGTGAAACAGAACTTGCGCTGGAGAA  
 QY 538 hrGluCysAenGlnCysLysAlaProLysProGlnGlyPheLeuProProPheProP  
 Db 1673 CAGAGTCAACCAAGTGTAAAGCCCAAGCCCTGAAGGCTTCTCCGACCCCTTCCGC  
 QY 558 rOProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyL  
 Db 1733 CCGGGGTGTATCTGGGAGAGAGTGCCTGTGTGACATGGGAGAGAGAGAGTGGC  
 QY 578 eMetAenPheArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyG  
 Db 1793 TCATGATCTGGGTGGTCCCGGAGGATGTTCAAGAGTGGCCGTGAGAGACAGAGTGG  
 QY 598 1yPheArgGlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyP  
 Db 1853 GCTTCCCTGTGTGGCCGGGAGATGACCGAGTGTGCTTGTGTGAGAGAACGAGTGGCC  
 QY 618 rOGlyGlyProProGlyProLeuMetGlnGlnMetGlyGlyArgArgGlyGlyArgGlyG  
 Db 1913 CTGGGGGGCCCCCTTGACCTTTATGAGAACAGATGGGAGAGAGAGAGAGCTGGAG  
 QY 638 1yProGlyLysMetAspLysGlyGlnHisArgGlnGlnArgArgAspArgProTyr  
 Db 1973 GACCTGGAATAATGATTAAGGAGAGCAACCTGACAGAGCCGACAGATCGGCCCTAC

RESULT 11  
 AAS62623/c  
 ID AAS62623 standard; cDNA; 2273 BP.  
 XX  
 AC AAS62623;  
 XX  
 DT 14-FEB-2002 (first entry)  
 DE  
 XX cDNA sequence #410 encoding novel human secreted protein.  
 XX  
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 KW immunosuppressive; antineumatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN NO2001.77291-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001MO-US010485.  
 XX  
 PR 06-APR-2000; 2000US-0195604P.



PA (GENY ) GENETICS INST INC.

XX Wong CG, Clark HF, Reichel K, Agostino MJ, Howes SH, Resnick RJ;

PI Gulukota K, Graham JR;

XX WPI; 2002-010900/01.

PT New polynucleotides encoding secreted proteins useful for treating e.g. asthma, HIV and Crohn's disease.

PS Claim 1, Page 291; 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences which encode human secreted proteins. The cDNA sequences have been derived from a variety of human tissues. The invention also provides a method for producing proteins from these polynucleotide sequences. The proteins are useful for identifying compounds that modulate their activity and production, and the cell is also useful for identifying compounds that modulate expression of the polynucleotide sequences encoding the secreted proteins. The sequences of the invention are useful for treating diseases such as hyperproliferative disorders (e.g. cancer), immune deficiency disorders (e.g. severe combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis). The polynucleotide sequences of the invention are also useful in gene therapy. AA62214-AA62838 represent the cDNA sequences of the invention that encode for novel human secreted proteins

XX Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,26e-111 Length: 2273  
Score: 3483.00 Matches: 636  
Percent Similarity: 97.10% Conservative: 0  
Best Local Similarity: 97.10% Mismatches: 1  
Query Match: 95.87% Indels: 18  
DB: 6 Gaps: 2

US-10-791-017A-2 (1-656) x AA62623 (1-2273)

QY 2 AAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAlaTyr 21  
DB 2263 GCGTCCAGCATTAAGTACTATAGCCAGCTGAGGCGAGCGGCTTACAGCTTAC 2264  
QY 22 ThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSerTyr 41  
DB 2203 ACCGCCAGCCACTCAAGATATGACAGACCAACCCAGCATATGGGCAACAAAGCTAT 2144  
QY 42 GlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyr 61  
DB 2143 GGAACCTATGAGACGCCAGTATGTCAGCTATACCCAGGCTCAAGACCACTGCAACTAT 2084  
QY 62 GlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrProThr 81  
DB 2083 GGGGAGACCGGCTATGCAACTTCTTATGACAGCTCCCACTGTTATACCTCAACT 2024  
QY 82 AlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThrThr 101  
DB 2023 GCCCCCGAGCATACAGCACCTGTCCAGGGGTATGCACTGGTGTATGATACCAACC 1964  
QY 102 ThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThrGln 121  
DB 1963 ACTGCTATAGTACCAACCAACCGAGCTCTTATGAGCTCAGCTGATATGAGCACTCAG 1904  
QY 122 ProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGlnAsp 141  
DB 1903 CCTGCTTATCAAGCTATAGGAGAGAGCAAGCACTGCACTTACAAAGCCGAGGAT 1844  
QY 142 GlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGlnPro 161  
DB 1843 GGAACAAGCCCACTGAGACTATGTCACCTCATCTATGACAGGGGGTTTACAAACAGCCC 1784

QY 162 SerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPromet 181  
DB 1783 AGCTTAGATATGAGACAGATTAAGTATATCCAGGTACTGGAGCTACCCCATG 1124  
QY 182 GlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThr 201  
DB 1723 CAGCCAGTCACTGCACTCCATCTTACCTCCACAGGATATTCCTTACAGCCGACT 1664  
QY 202 SerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyrGly 221  
DB 1663 AGTTATATCAAGAGCTTACTCTGACAGAACCTTATGGGCAACGACACTATGGA 1604  
QY 222 GlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyrPro 241  
DB 1603 CAGCAGATATGATATGATGATCAACAAACAGCTATGGGAGAGCTCCACTATGCCA 1544  
QY 242 ProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerTyr 261  
DB 1543 CCCCAACTGATCCATACAGCCCAAGCTCCCAAGTCAATATAGCCACAGAGCAGCTAC 1484  
QY 262 GlyGlnGlnSerSerPheArgGlnAspPheLysProSerSerMetGlyValTyrGlyGlnGlu 281  
DB 1483 GGGCAGCAGATTCATTCGACAGACCAACCCAGTAGAGTGGGTATTATGGGAGAG 1424  
QY 282 SerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArgGly 301  
DB 1423 TCTGAGAGATTTTCCGAGCAGAGAGAAACCGAGCATGATGCGCTGATACCGGGGC 1364  
QY 302 ArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGlyValArg 321  
DB 1363 AGGGGAAGGGGGATTGATCTGTGAGGCTGACAGAGGTGGGGGAGAGAGAGCGC 1304  
QY 322 GlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyPrometAsp 341  
DB 1303 GGTGAATGGGC--GCTGAGAGCCAGATGGCTTCAATAGCTCTGTGACCACTGAT 1247  
QY 342 GlnGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAsnSer 361  
DB 1246 GAAAGACAGATCTTATCTAGGCCCACTGTATGATCCAGTGAAGCTGTGACACAGT 1187  
QY 362 AlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspPheLeuAlaAspPhe 381  
DB 1186 GCAATTATATGACAGATTAATATGACAGTGTACTGTATATCTGCGAGACTCTTT 1127  
QY 382 LysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIleTyr 401  
DB 1126 AAGCAGGTGGGTGTGTAAGATGAACAGAGAACTGGGCAACATGATCCACATCTAC 1067  
QY 402 LeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysAspPro 421  
DB 1066 CTGAGCAGAGAAACAGGAAGGCCCAAGGAGTCCACAGTGTCTTATGAAGCCACCC 1007  
QY 422 ThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeuLys 441  
DB 1006 ACTGCCAAGGCTCCCTGATGATGTTGTGGGAAAGATTTTCAAGGGAGCAAACTTAA 947  
QY 442 ValSerLeuAlaArgLysLysProPrometAsnSerMetArgGlyGlyLysLeuProArg 461  
DB 946 GTCTCCCTGTCTGGAAGAAAGCTCCATATGAACAGTATGGGGGTGTGCAACCCGT 887  
QY 462 GlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyProGly 481  
DB 886 GAGGAGAGAGCATGCAACCACTCCGTGAGGTCCAGAGGCCAGAGAGGTCTCGGG 827  
QY 482 GlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArgGly 501  
DB 826 GACACCATGGGTGCACTAGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGA 767  
QY 502 ProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGlyAsp 521  
DB 766 CCCCGGGTTCGAGAGGAAACCTCTGAGAGGAGAAACCTCCAGCAGGAGCTGGAGAC 707  
QY 522 TrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThrGlnCysAsn 541

```

Db      706 TGGCAGTGTCCCATCCGGGTGTGGAACACGAACTTCGCTGAGAGACGACTGCAC 647
Qy      542 GlnCySelYbAlaProLybProGluGlyPheLeuProProPheProProGlyGly 561
Db      646 CAGGT-----GGT 638
Qy      562 AspArgGlyYbArgGlyYbProGlyYbMetArgGlyYbYbArgGlyYbLeuMetAspArg 581
Db      637 GATGTGGCAGAGGTGGCCCTGGTGGCATGGGGAGAGAGAGGTGGCTCATGATGCT 578
Qy      582 GlyYbProGlyYbMetPheArgGlyYbYbArgGlyYbYbAspArgGlyYbPheArgGly 601
Db      577 GGTGTCTCCGGTGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTTCGTGT 518
Qy      602 GlyYbArgGlyYbMetAspArgGlyYbYbPheGlyYbYbYbArgGlyYbYbProGlyYbPro 621
Db      517 GGGCGGGGCATGGACCGAGGTGGCTTGTGTGAGAGAGACAGAGTGGCCCTGGGGGGCCC 458
Qy      622 ProGlyProLeuMetGluGlnMetGlyYbYbArgGlyYbYbYbArgGlyYbYbProGlyYbYb 641
Db      457 CTGGACCTTGTGATGGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
Qy      642 MetAspLybGlyGlyYbYbYbArgGlyYbYbYbArgGlyYbYbProLyb 656
Db      397 ATGGATAAAGCGAGCACCGTCAGAGACCGAGAGATGGCCCTAC 353

RESULT 12
AAQ50646
ID      AAQ50646 standard; DNA; 3309 BP.
XX
AC      AAQ50646;
XX
DT      25-MAR-2003 (revised)
DT      26-MAY-1994 (first entry)
XX
DE      Human Ewe gene (genomic DNA).
XX
KW      chromosomal translocation; chimeric; chimeric; Ewing sarcoma; Ews gene;
KW      malignant melanoma; hum-fil-1;
KW      primitive peripheral neuroectodermal tumour; human chromosome 11;
KW      human chromosome 22; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FH      exon
FT      1. .37
FT      /+tag= a
FT      /number= 1
FT      /codon_start= 25. .27
FT      38. .96
FT      /+tag= b
FT      /number= 1
FT      /+tag= c
FT      /number= 2
FT      /+tag= d
FT      /number= 2
FT      /+tag= e
FT      /number= 3
FT      /+tag= f
FT      /number= 3
FT      /number= 3
FT      /+tag= g
FT      /number= 4
FT      /+tag= h
FT      /number= 4
FT      /+tag= i
FT      /number= 4
FT      /+tag= 1
FT      /number= 17

```

```

FT      /number= 5
FT      /+tag= j
FT      /number= 5
FT      /+tag= k
FT      /number= 6
FT      /+tag= l
FT      /number= 6
FT      /+tag= m
FT      /number= 7
FT      /+tag= n
FT      /number= 7
FT      /+tag= o
FT      /number= 8
FT      /+tag= p
FT      /number= 8
FT      /+tag= q
FT      /number= 9
FT      /+tag= r
FT      /number= 9
FT      /+tag= s
FT      /number= 10
FT      /+tag= t
FT      /number= 10
FT      /+tag= u
FT      /number= 11
FT      /+tag= v
FT      /number= 11
FT      /+tag= w
FT      /number= 12
FT      /+tag= x
FT      /number= 12
FT      /+tag= y
FT      /number= 13
FT      /+tag= z
FT      /number= 13
FT      /+tag= aa
FT      /number= 14
FT      /+tag= ab
FT      /number= 14
FT      /+tag= ac
FT      /number= 15
FT      /+tag= ad
FT      /number= 15
FT      /+tag= ae
FT      /number= 16
FT      /+tag= af
FT      /number= 16
FT      /+tag= ag
FT      /number= 17

```

FT polyA\_signal 3102..3107  
 FT polyA\_signal /\*\*tag= ah  
 FT polyA\_signal 3289..3294  
 FT polyA\_signal /\*\*tag= al  
 XX MO9323549-A2.  
 XX 25-NOV-1993.  
 XX 19-MAY-1993; 93MO-FR000494.  
 XX 20-MAY-1992; 92FR-00006123.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Aurias A, Delattre O, Desmaze C, Meiot T, Peter M, Plougaetel B,  
 PI Thomas G, Zucman J,  
 XX WPI; 1993-386580/48.  
 DR P-PDB; AAR44555.  
 XX  
 PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion  
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
 XX  
 PS Disclosure; Fig 6 and Fig 12; 123pp; French.  
 CC The intron-exon junctions of the human Ews gene and the Hum-F11-1 gene  
 CC have been sequenced (see AAQ50646 and AAQ50662, respectively). The  
 CC different fusion products which could be formed by fusing exons from the  
 CC two genes, as happens after specific chromosomal translocations, can be  
 CC predicted (see AAQ50671-Q50678). The sequences at fusion junctions of  
 CC other observed translocations are given in AAQ50679-Q50683). NOTE: the  
 CC 3309 residue sequence AAQ50646 does not appear in the specification; it  
 CC is a combination of the Ews cDNA sequence (Fig 6, AAQ50643) and the  
 CC individual intron sequences with their intron-exon junctions (Fig 12).  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 3309 BP; 822 A; 775 C; 857 G; 855 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1.71e-105 Length: 3309  
 Score: 3315.00 Matches: 652  
 Percent Similarity: 68.49% Conservative: 0  
 Best Local Similarity: 68.49% Mismatches: 300  
 Query Match: 91.25% Indels: 15  
 DB: Gaps: 15  
 US-10-791-017a-2 (1-656) x AAQ50646 (1-3309)  
 QY 5 Aspyrterthrrtyrserglnalalaglninglly----- 17  
 DB 96 GATTACAGTACTTATGACCAAGTCGACGAGGAGGTAAGTCACTTTTAAACG 155  
 QY 18 -----Tyrseralatyrrhalagln 24  
 DB 156 TATTTTGTTCAGTATGATGATTAATCTTTTGACGACGACGCTTACACGCCCCAG 215  
 QY 25 Prothrglnglytyralaglnthr----- 33  
 DB 216 CCCACTCAAGATATGACACACCAAGGTAATCTTTAAATAATACATGATGCTGC 275  
 QY 34 -----Glnalatyrglynglnglinsertyrglythrttyr 44  
 DB 276 GTTCTTCATTCGGTTTTTTTTTGGAGGACGATATGGCAAAACCTATGAACTAT 335  
 QY 45 GlyglnProthraapValserTyrrhrglnalaglnThrThralatrrTyrglyglnThr 64  
 DB 336 GGAAGGCCCACTATGTCAGCTATACCAAGGCTACAGCACTGCAACTATGAGCAAGC 395  
 QY 65 AlaTyralatrrserTyrglyglnProth----- 75  
 DB 396 GCGTATGCAACTTCTTATGACAGCCTCCACTGTGTAAGCGCTGAGAGATTTT 455

QY 76 -----GlyTyrrhThrrProthralaProglIn 84  
 DB 456 GGGTGAATCTGATGACGCTCCCTTTGGCTTACAGTATACCTCACTGCCCCCAG 515  
 QY 85 AlaTyrrserglnProvalaglnglytyrglythrglyalaTyrrapThrrThralatrr 104  
 DB 516 GATACAGCAGCTGTCCAGGGGTATGGCACTGGGTATATATACCACTGCTACCA 575  
 QY 105 ValThrrThrglnalaserTyrralalaglnserAlatyrGlythrglnProalatyrr 124  
 DB 576 GTCACCAACCAAGGCTCTCTATGCACTGACGCTGATGCACTGCACTGCTTAT 635  
 QY 125 ProAlaTyrglynglnglnProAlaThralaProTh----- 137  
 DB 636 CAGCCTATGGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 685  
 QY 138 -----ArgProglInapGlyAenlyg 144  
 DB 696 AATGCGTCACTTATATTTATTTATTTCTCTCTTACGACGAGATGGAACAAG 755  
 QY 145 ProThrglnThrrserglnProglInserSerThrglytyrrhnglnProserleugly 164  
 DB 756 CCCACTGAGACTATGACCACTCAATCTTACACAGGGGTTAAACAGCCACCTTACGA 815  
 QY 165 TyrglyGlnSerapnTyrrserTyrrProglInvalProglSerTyrrPrometGlnProval 184  
 DB 816 TATGACAGAGTACTACAGTTATCCCAAGTACTGGAGCTACCCCATGACGACGATC 875  
 QY 185 ThrAlaProProserTyrrProTh----- 193  
 DB 876 ACTGACCTTCATCTTACCTCTCAAGTCACTTTTGGCAAAACAAT 935  
 QY 194 -----SerTyrrserThrglnProthrrserTyrrap 204  
 DB 936 TTTTTCCT 955  
 QY 205 GlnSerSerTyrrserglnInaentThryrglyglnProserSerTyrglyglnInser 224  
 DB 996 CAGACGATTTACTCTACAGCAACCACTTATGGCAACGACGACGACGACGACGACG 1055  
 QY 225 SerTyrglyglnInserSerTyrglyglnInProProthrrserTyrrProglInthr 244  
 DB 1056 AGCTATGCTCAACAAAGCAGTATGGGACGACCTCCCACTGATTAACCAACCCAACT 1115  
 QY 245 GlySerTyrrserglnalaproserglnTyrrserglnInserSerSerTyrglyglnIn 264  
 DB 1116 GATCTTACAGCCAAAGCTCCAACTCAATATACCAAGACGACGACGACGACGACG 1175  
 QY 265 Ser----- 265  
 DB 1176 AGTGAAGTGTGAAGAGAAAACCAATAAGCATGAGCTTACATGCTCTTCTCA 1235  
 QY 266 ---SerPheArgGlnapnIapProserSerMetGlyValTyrglynglnglserGly 284  
 DB 1236 GGTTCATTCGACAGACCAACCCAGATGATGGGTATATGGGACGAGTCTGAGAGA 1295  
 QY 285 PheSerGlyProglInvaluaenargserMetSerGlyProapbnsmnrglyIaargGlyArg 304  
 DB 1296 TTTTCCGAGACAGAGAAACCGAGACATGAGGCTCGATTAACCGGGGACAGGGAAGA 1355  
 QY 305 GlyGlyPheapbnrglyGlyMetSerArgGlyIaargGlyIaargGlyIaargGlyMet 324  
 DB 1356 GGGGAGTTTATCTGTGAGGACATGACAGAGTGGGCGGGAGAGAGACCGGTGAATG 1415  
 QY 325 Gly----- 325  
 DB 1416 GGTATAGAGCAACCTTTCTCTTTTACCTAAAGGCTTATCTGTTATGCCCC 1475  
 QY 326 ---SerAlaGlyGluArgGlyGlyPheapbnIapProgl----- 337  
 DB 1476 AGCAGGCGTGAAGAGGAGGCTTCAATAGCCTGTGTGTAAGTTTGAATTAAC 1535

```

Qy 338 -----GlyPrometAspGlyPro 344
Db 1536 ATAGATATGATATATTTATATGATCTTCTGTTGGAGGAGCCATGATGAAGACCA 1595
Qy 345 AspLeuAspLeu----- 348
Db 1596 GATCTTATCTAGAGTAAGTGAATTCCTAGTTGTCCTTCATATATTCCTCTGTCCTG 1655
Qy 349 -----GlyProProValAspProAspGlyAspSerAspAsnSerAlaIleTyr 364
Db 1656 TTCTCTCTGAAGAGCCCATCTGATATCCAGTATACACTGTGCAACAGGCAATTTAT 1715
Qy 365 ValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPheLeuGlyCys 384
Db 1716 GTTCAAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAAGCTTCTTAAAGCAAGTGT 1775
Qy 385 GlyValValIys----- 388
Db 1776 GGGGTTGTTAAGTCAGTAAAGCATTAACAGTCATCTGGCTCATGCTTAACATATGCTA 1835
Qy 389 -----MetAsnLysArgThrGlyGlnProMetIleHisIleTyrLeuAspLys 404
Db 1836 TTCTTTCTAGATGAACAAGAGACTGGGCAACCATGATCCATCTACCTGAGACAG 1895
Qy 405 GluThrGlyLysProLysGlyAspAlaThrValSerTyrGlyAspProProThrAlaLys 424
Db 1896 GAAACAGGAAAGCCCAAGGAGATGCAAGTGTCTATGAAGACCAACCATGCGCAAG 1955
Qy 425 AlaAlaValGlyTyrPheAsp----- 431
Db 1956 GCTGCGGTGAATGCTTTGATGTGAGATGACTCAGTGGCATTTCTAATCTAGTATTTG 2015
Qy 432 -----GlyLysAspPheGlnGlySerLysLeuLysValSerLeu 444
Db 2016 ATGTTCTGTGTCTTGTTCAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTT 2075
Qy 445 AlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProProArgGlyLysArg 464
Db 2076 GCTTCGAAAGAGCCTCCAAATGACATGATGGGGGTGCTGCAACCCCGAGGAGCA 2135
Qy 465 GlyMetProProProLeuArgGly----- 472
Db 2136 GGCATGCAACCAACCTCCGTGAGGACTTTACTGAGCTCCATGTTGATTAATTTG 2195
Qy 473 -----GlyProGlyLysProGlyLysProGlyLysProMet 484
Db 2196 CTGTTCTTGTGTTCTTGTGTGATGATGAGGAGGCTCTGAGGAGGAGCCCATG 2255
Qy 485 GlyArgMetGlyLysLysArgGlyLysAspArgGlyLysPheProProAspGlyProArgGly 504
Db 2256 GGTTCGATGGAGGAGCCGTGAGAGATGAGAGGCTTCCCTCCAAAGAGACCCCGGGGT 2315
Qy 505 SerArgGlyAsnProSerGlyLysLysValGlnHisArgAlaGlyAspTyrGlnCys 524
Db 2316 TCCGAGGAGAACCCCTCTGAGAGAGAAAGTCCAGACAGAGTGAAGTGGCACTGT 2375
Qy 525 ProAsnPro----- 527
Db 2376 CCCAATCCGATATCTTGTCTGGGAAATATGATACCTGATTTCTGCTGATGATAT 2435
Qy 528 -----GlyCysGlyAsnGlnAsnPheAlaTyrArgThrGlyCysAsnGlnCysLys 544
Db 2436 GATTCAGAGGGGTGTGAAACCAAGACTTCCCTGAGAAACAGAGTGAACAGGTGAAG 2495
Qy 545 AlaProLysProGlnGlyPheLeuProProPheProProPro----- 559
Db 2496 GCCCAAGAGCTGAAGCTTCTCTCCGCAACCTTTCGCGCCCGGGGTAGAGTGAAGTTT 2555
Qy 560 -----GlyGlyAspArgGly 564
Db 2556 CATGAGTGTCCCTCAACCTGTTCCGCTGATTTCTCACTTGAAGTGTGATCGGAG 2615
Qy 565 ArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeuMetAspArgGlyLysPro 584

```

```

Db 2616 AGAGTGGCCCTGCTGATGCGGGAGAGAGAGTGGCTCATGATCGTGTGTCCTC 2675
Qy 585 GlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLysPheArgGlyLysArgGly 604
Db 2676 GGTGAATGTTTCAGAGGTGCGCCGTGTGAGAGACAGAGTGGCTTCCGTGGTGGCGGGGC 2735
Qy 605 MetAspArgGlyLysPheGlyLysGlyLysArgArgGlyLysProGlyLysProProGlyPro 624
Db 2736 ATGACCGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGGGGGGCCCTGGAAGT 2795
Qy 625 LeuMetGlnGlnMetGlyLysArgArgGlyLysArgGlyLysProGlyLysMetAspLys 644
Db 2796 TTGATGAAACAGATGGAGAGAGAGAGAGACGTGAGAGACCTGGAATAATGATTAAG 2855
Qy 644 ----- 644
Db 2856 TAAGTCTGGTGAAGAACAGCTGTGGGCTCTAACGGAAGGCCCTTTACCTTGAGA 2915
Qy 645 GlyLysHisArgGlnGlnArgArgAspArgProTyr 656
Db 2916 GCGAGCACCGTCAAGAGCGCAGAGATCGGCCCTTAC 2951

RESULT 13
ADA53506
ID ADA53506 standard, cDNA, 1988 BP.
XX
AC ADA53506;
XX
DT 20-NOV-2003 (first entry)
DE Human coding sequence, SEQ ID 1074.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX Inflammatory disease; osteoporosis; neurological disease; gene; ss.
OS Homo sapiens.
XX
PN BP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
DR WPI; 2003-395539/38.
DR P-PDB; ADA55145.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1074; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

```

Alignment Scores:



XX 04-NOV-2004 (first entry)  
 XX Full length human cDNA useful for treating neurological disease Seq 952.  
 DE  
 XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;  
 XX osteoporosis; neurological disease; Alzheimer's disease;  
 XX Parkinson's disease; dementia; short memory; cancer;  
 XX sense or motor function; emotional reaction; fear response; panic;  
 XX osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;  
 XX tranquilizer.  
 OS Homo sapiens.  
 XX  
 XX EP1447413-A2.  
 XX  
 XX 18-AUG-2004.  
 XX  
 XX 12-FEB-2004; 2004EP-00003145.  
 XX  
 XX 14-FEB-2003; 2003JP-00102207.  
 XX 09-MAY-2003; 2003JP-00131452.  
 XX  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 XX Ilogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,  
 XX Wakamatsu A, Ishii S, Nagai K, Irie R,  
 XX WPI; 2004-583265/57.  
 XX P-PSDB; ADR09402.  
 XX  
 XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
 XX Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.  
 XX  
 XX Claim 1; SEQ ID NO 952; 2686bp; English.  
 XX  
 XX This invention relates to novel, isolated full length human cDNA  
 XX molecules and the encoded proteins thereof. Specifically, it refers to  
 XX cDNA clones obtained by an oligo-capping method, where none of these  
 XX clones are identical to any known human mRNAs. The present invention  
 XX describes an immunoassay to identify agonists and antagonists, as well as  
 XX antibodies, antisense molecules and siRNAs that can all be used to bind  
 XX to and modulate expression of the cDNA molecules. As such, these  
 XX molecules are useful for diagnostic markers or therapeutic targets for  
 XX the various diseases or morbid states. In particular, they are useful in  
 XX gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
 XX disease, Parkinson's disease, dementia, short memory and various cancers,  
 XX as well as for maintaining equilibrium of sense or motor function, and  
 XX for treating emotional reaction, fear response and panic. Accordingly,  
 XX they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
 XX cyostatic and tranquilizer activities. This polynucleotide is a full  
 XX length human cDNA sequence of the invention. NOTE: This sequence is not  
 XX given in the sequence listing of the specification but can be obtained on  
 XX CD-ROM from the European Patent Office, Vienna Sub-office.  
 XX  
 XX Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 4, 71e-102 Length: 2026  
 XX Score: 3210.00 Matches: 590  
 XX Percent Similarity: 90.40% Conservative: 3  
 XX Best Local Similarity: 89.94% Mismatches: 11  
 XX Query Match: 88.36% Indels: 52  
 XX DB: 13 Gaps: 4  
 XX  
 XX US-10-791-017A-2 (1-656) X ADR07446 (1-2026)  
 QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlnTyrSerAla 20  
 DB 38 ATGGCGTCCAGCATTAAGTACTATAGCCAGCTGCGACGCGCGGCTACAGTCT 97  
 QY 21 TTTThAlaGlnProThrgInGlnTyrAlaGlnThThrGlnAlaTyrGlnGlnSer 40

DB 98 TACACGCCACGCCACTCAAGATATGACACAGCACCCAGGATATGGCAACAAC 157  
 QY 41 TTTGlyThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 DB 158 TATGAACTATATGAGACAGCCCACTGATGCTACATATCCAGGCTCAGACCACTGCAAC 217  
 QY 61 TTTGlyGlnThrAlaTyrAlaThrSerTyrGlnGlnProThrGlnTyrThrThrPro 80  
 DB 218 TATGGGAGACCGCTATATGAACTCTTATGAGACGCTCCACTGGTTTATCACTCCA 277  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnTyrGlnThrGlnAlaTyrAspThr 100  
 DB 278 TCTACCTCTCTAC-----TACTCTCAGACAGCCAGCCACTGACCTACAGACCCGAG 304  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlnThr 120  
 DB 305 ACACAGCCGACT-----AGTTATGATCAGAGAGCT----- 334  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnGlnProAlaAlaThrAlaProThrAspProGln 140  
 DB 335 -----TACTCTCAGACAGCCAGCCACTGACCTACAGACCCGAG 376  
 QY 141 AspGlyAsnLysProThrgInThrSerGlnProGlnSerSerThrgInGlyTyrAsnGln 160  
 DB 377 GATGAAACAGCCCACTGAGACTAGTCAACCTCACTGACAGGGGTTTACACAG 436  
 QY 161 ProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 437 CCCAGCTTATGATATGAGACAGAGTAACACTACAGTTATCCAGGATACCTGAGAGTACCCC 496  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrgInPro 200  
 DB 497 ATGACAGCACTGACTGACCTGACCTCACTCACTCTTCACTCACT----- 535  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlnGlnProSerSerTyr 220  
 DB 536 -----AGCTAT 541  
 QY 221 GlyGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnGlnProProThrSerTyr 240  
 DB 542 GACAGAGAGTATGATATGATCAACAAGCAGCTATGAGCAGCAGCTCCCACTAGTTAC 601  
 QY 241 ProProGlnThrThrGlnSerTyrSerGlnAlaProSerGlnThrSerGlnGlnSerSer 260  
 DB 602 CCACCCCAACTGATCTCAACACCAAGCTCCAAAGTCMAATATGACCAACAGACAGCAGC 661  
 QY 261 TTTGlyGlnGlnSerSerPheAArgGlnAspHisProSerSerMetGlyValTyrGlnGln 280  
 DB 662 TACGGGACAGAGCTTCACTTCCAGACAGACACCCCACTGATGGTGTATTATGGCAG 721  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 722 GACTGTGAGAGATTTTCCGAGACAGAGAGAACCGAGCATGATGGCCCTGATTAACCG 781  
 QY 301 GlyArgGlyValArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGly 320  
 DB 782 GCGAGGGGAGAGAGAGGGGATTTGATCGTGGAGCATAGAGAGTGGCGGGAGAGAGGA 841  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 DB 842 CGCGGTGGAATGGCAGCGCTGAGAGAGAGAGGTGGCTTCAATTAAGCTGTGTGAGCCATG 901  
 QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspAsn 360  
 DB 902 GATGAAGAGACAGATCTTATCTAGGCCCACTGTGATCAATGATGAAGACTCTGACAAAC 961  
 QY 361 SerAlaIleTyrTalaGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 962 AGTGAATTTATATACAGAGATTAATGAACAGTGTACTTATATATATCTGAGAGACTTC 1021  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1022 TTTTACAGTGTGGGTGTTTAAATGATGACAGAGAACTGGGCAACCATGATCATCATC 1081







**This Page Blank (usph)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 22:00:21, Search time 291.042 Seconds  
(without alignments)  
3688.116 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633

Sequence: 1 MASTDSTYSQAANAQGYSA.....GGPCKDKGHRGRRDRPY 656

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Fgapop 10.0, Fgapext 0.5	
Dgapop 6.0, Dgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10791017/runat.17022005.125808.22107/app.query.fasta\_1.1358  
-DB=Issued Patents NA -OEMT=fastap -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -INITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt  
-LIST=45 -DOCALL=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=/cgn2\_1/USPTO.spool\_p/cgn2\_1.141@runat.17022005.125808.22107 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOR=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/lna/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/lna/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2371	2	US-08-343-443B-1
2	3233	89.0	6002	4	US-09-949-016-13636
3	3178.5	87.5	1785	4	US-09-949-016-5043
4	3178.5	87.5	1785	4	US-09-949-016-5044
5	2826.5	77.8	1783	4	US-09-949-016-1954
6	1420	39.1	2432	1	US-08-437-027-18
7	1273.5	35.1	1939	4	US-09-919-039-322
8	958	26.4	35784	4	US-09-949-016-16785
9	958	26.4	35784	4	US-09-949-016-16786
10	676	18.6	454	4	US-09-513-999C-1657
11	609	16.8	411	4	US-09-621-976-13361
12	566	15.6	601	4	US-09-949-016-66382

13	486	13.4	1682	4	US-09-320-132-82	Sequence 82, Appl
14	484	13.3	1954	2	US-08-343-443B-106	Sequence 106, App
15	484	13.3	1684	4	US-09-919-039-323	Sequence 323, App
16	471	13.0	333	4	US-09-513-999C-12062	Sequence 12062, A
17	405	11.1	601	4	US-09-949-016-176641	Sequence 176641,
18	405	11.1	601	4	US-09-949-016-176664	Sequence 176664,
19	397.5	10.9	450	4	US-09-370-838-145	Sequence 145, App
20	397.5	10.9	450	4	US-09-854-133-145	Sequence 145, App
21	386.5	10.6	3147	2	US-08-781-802-7	Sequence 7, Appl
22	386.5	10.6	3147	3	US-08-694-078-7	Sequence 7, Appl
23	386.5	10.6	3147	3	US-09-058-260-7	Sequence 7, Appl
24	368.5	10.1	3600	1	US-08-537-002A-5	Sequence 5, Appl
25	368.5	10.1	3600	3	US-08-863-010-5	Sequence 5, Appl
26	368.5	10.1	3600	3	US-09-024-429-5	Sequence 5, Appl
27	360.5	9.9	2830	2	US-09-010-928B-1	Sequence 1, Appl
28	353.5	9.7	6109	4	US-09-795-061-1	Sequence 1, Appl
29	351.5	9.7	1936	3	US-09-249-585A-4	Sequence 4, Appl
30	351.5	9.7	1931	2	US-09-130-114-2	Sequence 2, Appl
31	347	9.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
32	345	9.5	4411529	3	US-09-103-840A-1	Sequence 1, Appl
33	344	9.5	2082	4	US-09-818-780-67	Sequence 67, Appl
34	343	9.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
35	341.5	9.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
36	341	9.4	2824	2	US-09-010-928B-3	Sequence 3, Appl
37	337.5	9.3	4409	4	US-09-331-347C-22	Sequence 22, Appl
38	331	9.1	6200	4	US-09-795-061-3	Sequence 3, Appl
39	330	9.1	6158	4	US-09-919-497-6	Sequence 6, Appl
40	329.5	9.1	3171	4	US-09-169-768-15	Sequence 15, Appl
41	329.5	9.1	3181	1	US-08-655-086-1	Sequence 1, Appl
42	329.5	9.1	3349	4	US-09-169-768-13	Sequence 13, Appl
43	329.5	9.1	3541	4	US-09-169-768-9	Sequence 9, Appl
44	329	9.1	2493	1	US-07-977-434-5	Sequence 5, Appl
45	329	9.1	2493	1	US-08-458-819-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploongastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zuchman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343, 443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR93/00494  
 FILING DATE: 19-MAY-1993  
 PRIOR APPLICATION DATA: FR 92/06123  
 APPLICATION NUMBER: 20-MAY-1992  
 FILING DATE: 20-MAY-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Weiser, Gerard J.  
 REGISTRATION NUMBER: 19,763  
 REFERENCE/DOCKET NUMBER: 989,6121P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-875-8383  
 TELEFAX: 215-875-8394  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2371 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 25..1992  
 US-08-343-443B-1

Alignment Scores:  
 Pred. No.: 4.3e-186 Length: 2371  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-10-791-017a-2 (1-656) x US-08-343-443B-1 (1-2371)

QY 1 MetAlaSerThrAspIYrSerThrTYrSerGlnAlaAlaAlaGlnGlnGlyTYrSerAla 20  
 DB 25 ATGGCGTCCACGAGATTACAGTACCTATAGCCAGAGCTCAGCGGCGCTACAGTCT 84  
 QY 21 TTTTThrAlaGlnProThrGlnGlyTYrAlaGlnThrGlnAlaTYrGlyGlnGlnSer 40  
 DB 85 TACACCGCCGACCTCAGATGATGACAGACCAACCGAGCATATGGCAACCAAGC 144  
 QY 41 TYrGlyThrTYrGlyGlnProThrAspValSerTYrThrGlnAlaGlnThrAlaThr 60  
 DB 145 TATGGAACTATGACAGCCCACTGATGTACGCTATACCCAGGCTCAGACCACTGCAAC 204  
 QY 61 TYrGlyGlnThrAlaTYrAlaThrSerTYrGlyGlnProProThrGlyTYrThrThrPro 80  
 DB 205 TATGGGACGACCGGCTATGCAACTTCTTATGACAGGCTCCCACTGTTTACTACTCCA 264  
 QY 81 ThrAlaProGlnAlaTYrSerGlnProValGlnGlyTYrGlyThrGlyAlaTYrAspThr 100  
 DB 265 ACTGCCCCCGGAGCATACAGCCAGCTGTCCAGGGGATGAGCATGGTCTTATGATACC 324  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTYrAlaAlaGlnSerAlaTYrGlyThr 120  
 DB 325 ACCACTGCTATAGTACACACCAACCGGCTCTTATGAGCTCAGTCTGATATGGCACT 384  
 QY 121 GlnProAlaTYrProAlaTYrGlyGlnGlnProAlaAlaAlaThrAlaProThrAspProGln 140  
 DB 385 CAGCTCTTATCCAGGCTATGAGGACAGACGACCACTGACCTTACCAAGACCCGAG 444  
 QY 141 AspGlyAlaAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTYrAsnGln 160  
 DB 445 GATGAAACCAAGCCCACTGAGACTAGTCACTCACTATGACCAAGGGGATTACACAG 504  
 QY 161 ProSerLeuGlyTYrGlyGlnSerAsnLysSerTYrProGlnValProGlySerTYrPro 180  
 DB 505 CCCAGCTTAGATATGACAGAGTACACTACAGTTATCCCAAGTACTCGGAGAGTACCCC 564  
 QY 181 MetGlnProValThrAlaProProSerTYrProProThrSerTYrSerSerThrGlnPro 200  
 DB 565 ATGACAGCACTGACCTGACCTTCACTCACTCTCTACAGCTATTCCTCTACACAGCCG 624

QY 201 ThrSerTYrAspGlnSerSerTYrSerGlnGlnAlaThrTYrGlyGlnProSerSerTYr 220  
 DB 625 ACTAGTTATGATCAGAGCACTTACTCTCAGCAAGAACCTTATGGGACAGCACTAT 664  
 QY 221 GlyGlnGlnSerSerTYrGlyGlnGlnSerSerTYrGlyGlnGlnProProThrSerTYr 240  
 DB 685 GACACAGCAGATGATGATGCTCAACAAACACTATGGCAGACCTCCACTATGATAC 744  
 QY 241 ProProGlnThrGlySerTYrSerGlnAlaProSerGlnThrSerGlnGlnSerSer 260  
 DB 745 CCACCCCAACTGATCTTACAGCCAGGCTCCAAAGTCAATATGACCACAGACAGCAG 804  
 QY 261 TYrGlyGlnGlnSerSerPheAspGlnAspHisProSerSerMetGlyValTYrGlyGln 280  
 DB 805 TACGGGACACAGAGTTTCACTTCGACAGACCAACCCAGATGATGGTGTATTATGGGACG 864  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 865 GAGTCTGAGAGATTTTCCGAGCAGAGAGAAACCGAGCATGATGCGCTGTAAACCGG 924  
 QY 301 GlyArgGlyArgGlyValPheAspArgGlyGlyMetSerArgGlyValArgGlyGly 320  
 DB 925 GGCAGGGGAGAGGGGATTTGATCGTGAAGCATGACAGAGTGGCGGGAGAGAG 984  
 QY 321 ArgGlyGlyMetGlySerAlaGlyAlaGlyGlyGlyPheAsnLysProGlyGlyPromet 340  
 DB 985 CCGGCTGGAATGGGACGCGCTGGAAGGACAGTGGCTTCAATAGCTGTGGTGAACCAAT 1044  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360  
 DB 1045 GATGAAGGACCAAGATCTGATCAGGCGCTCTGATGATCCAGATGACCTCTGCAAC 1104  
 QY 361 SerAlaIleTYrValGlnGlyLeuAspSerValThrLeuAspLeuAlaAspPhe 380  
 DB 1105 AGTGCAATTTATGTAACAAGATTTAAATGAAGTGTACTTCAATATGATCTGGCAGACTTC 1164  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1165 TTTTACAGATGTGGGTTGTTAAGATGAACAAGAGAACCTGGGCAACCATGATCCACATC 1224  
 QY 401 TYrLeuAspLysGlyuThrGlyLysProLysGlyAspAlaThrValSerTYrGluAspPro 420  
 DB 1225 TACCTGACAGAGAAACAGAAAGCCCAAGGAGTGCACAGTGTCTTATGAAAGACCCA 1284  
 QY 421 ProThrAlaLysAlaAlaValGluTProPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1285 CCCACTGCCAAGGCTCCGGAATGTTGATGGGAAAGATTTTCAAGGAGACCAAACTT 1344  
 QY 441 LysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyLysLeuProPro 460  
 DB 1345 AATGCTCTCTTGTCTCGGAAGAACCTCCAAATGAACATATGGGGGTGTGCCACCC 1404  
 QY 461 ArgGlnGlyArgGlyMetProProPheLysArgGlyGlyProGlyGlyProGlyGlyPro 480  
 DB 1405 CGTGAAGGACAGAGGATGACCAACCACTCCGTGAGGTGCCAGAGAGCCCAAGAGGTCT 1464  
 QY 481 GlyLysProMetGlyValMetGlyValArgGlyGlyAspArgGlyGlyLysPheProArg 500  
 DB 1465 GGGGAGACCATGGGTGTGATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGA 1524  
 QY 501 GlyProArgGlySerArgLysAsnProSerGlyGlyLysValGlnAlaAlaArgAlaGly 520  
 DB 1525 GAGCCCCGGGTTCCGAGAGGAACCCCTCTGAGAGAGGAACCTCCAGACCGAGGTGGA 1584  
 QY 521 AspTYrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThrGluCys 540  
 DB 1585 GACTGGCAGATGCCATCCGAGGTGTGGAACCAAACTTCCCTCGAGAAACAGGTGC 1644  
 QY 541 AsnGlnCysValAlaProLysProGlyGlyPheLeuProProProPheProProGly 560  
 DB 1645 AACCAAGTGAAGGCGCCCAAGCCTGAAGGCTTCTCCGCAACCTTCCGCGCGGGGT 1704  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580

Db 1705 GGTGATGTGGCAGAGGTGGCCCTGTGTGCATGTGGGGAGAAAGAGTGGCTTATGAT 1764  
Qy 581 ATGGLYGLYProGlyGlyMetPheArgGlyGlyValArgGlyGlyAspArgGlyGlyPheArg 600  
Db 1765 CGTGTGTGTCCCGGTGGAATGTTTCAAGAGTGGCCGTGTGGAACAAGGTGGCTTCGT 1824  
Qy 601 GLYGlyValArgGlyMetAspArgGlyGlyPheGlyGlyValArgGlyGlyProGlyGly 620  
Db 1825 GGTGGCCGGGAGAGAGCCAGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCTTGGGGGG 1884  
Qy 621 ProProGlyProLeuMetGlnGlnMetGlyGlyValArgArgGlyGlyProGly 640  
Db 1885 CCCCCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGACCTGAGAGACTGGA 1944  
Qy 641 LysMetAspLysGlyGlyLysValArgGlnGlnArgAspArgProTyr 656  
Db 1945 AAAATGATTAAGCGAGACCGCTCAGAGCGCAGAGATCGCCCTAC 1992

RESULT 2  
US-09-949-016-13696  
/ Sequence 13696, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTUR, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
/ FILE REFERENCE: CL001307  
/ CURRENT APPLICATION NUMBER: US/09/949, 016  
/ PRIOR FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20  
/ PRIOR APPLICATION NUMBER: 60/237,768  
/ PRIOR FILING DATE: 2000-10-03  
/ PRIOR APPLICATION NUMBER: 60/231,498  
/ PRIOR FILING DATE: 2000-09-08  
/ NUMBER OF SEQ. ID NOS: 207012  
/ SOFTWARE: FASTSEQ for Windows Version 4.0  
/ SEQ ID NO 13696  
/ LENGTH: 6002  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-949-016-13696

Alignment Scores:  
Pred. No.: 2,398-164 Length: 6002  
Score: 3233.00 Matches: 601  
Percent Similarity: 93.60% Conservative: 13  
Best Local Similarity: 91.62% Mismatches: 42  
Query Match: 88.99% Indels: 2  
DB: Gaps: 0

US-10-791-017a-2 (1-656) x US-09-949-016-13696 (1-6002)

Qy 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
Db 2034 ATGGCATTCATGAGATTACATTAACCTTAAGCCAGCTCAGCGCAGGCTCAAGTCT 2093  
Qy 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaGlnThrAlaThr 40  
Db 2094 TACACACCCAGCCCACTCAAGATATGACACACCCAGGAGATATGGGACACAAAGC 2153  
Qy 41 TyrGlyThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60  
Db 2154 TATGAAACCTATGAGACGCCCATGTGATGCTAGCTATACCCAGGCTCAGACCATGCAATC 2213  
Qy 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrPro 80  
Db 2214 TATGGGAGACCGCTATGCAACTTCTTATGAGACAGCTCCCATGTGTATTAATCACTCA 2273  
Qy 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100  
Db 2274 ACTGCCCCCAGGATACAGCCAGCTGTCTCAGGGGATATGGCACTGTGCTTATGATACC 2333

Qy 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
Db 2334 ACCACGCTACAGTACACACACCAGCCGCTCTTATGACCTCAATGTGATTAAGGACT 2393  
Qy 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
Db 2394 CAGCTCTTATTCAGCTTATGGGACAGCCAGTACCCACTGACCTTACAGACCCGAG 2453  
Qy 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
Db 2454 AATGAAACAGCCCACTGAGACTATGACACTTCAATCTGACAGAGGGGTTCACACAG 2513  
Qy 161 ProSerLeuGlyTyrGlyGlnSerAspTyrSerTyrProGlnValProGlySerTyrPro 180  
Db 2514 CCAAGCTTAAATGAGACAGAGTAACTGACATTAATCCAGGATCCTGGAAGCTACTCC 2573  
Qy 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
Db 2574 ATGACGCCAGTACCGCAGCTTCAATCTTACCTTCAAGCTATTCCTTACACAGCCA 2633  
Qy 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
Db 2634 ACTAGTTATGATCAGAGCACTTACTCTCAGACAGAACCTATGGAAACCGAGAGCTAT 2693  
Qy 221 GLYGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
Db 2694 GACAGCAGAGTATGCTATGCTCAACAAAGCAGTATGGGAGCTGCTCCACTAGTTAC 2753  
Qy 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnThrSerGlnGlnSerSerSer 260  
Db 2754 CACACCCAACTGATCTACAGCAAGCTCCAAATGATGATGATCAACAGAGAGAGAGC 2813  
Qy 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
Db 2814 TACGGGAGCAGAGTTCATTCTGACAGAGACACCCAGTGAGCATGGGTGTTATGGCAG 2873  
Qy 281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300  
Db 2874 GAGTCTGAGAGATTTTCCGACATGAGAGAACCCGAGCATGATGAGTCCCTGATTAATCGG 2933  
Qy 301 GLYArgGlyValArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGlyGly 320  
Db 2934 GGCAGGGGAAAGAGGGGATTTGATCGGAGGATGAGCAGAGCTGGCGGGAGAGAGGA 2993  
Qy 321 ArgGlyGlyMetGlySerAlaGlyGlyArgGlyGlyPheAsnLysProGlyGlyProMet 340  
Db 2994 TGGGTTGGAATGGGACGCGCTGAGAGCAAGTTGCTTCAATTAAGCTGTGAGACCATG 3053  
Qy 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
Db 3054 GATGAAGACCAAGATCTTGATCTAGGCCCACTGTGATCAATGATGAAGACTGACACAC 3113  
Qy 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
Db 3114 AGTGAATTTATTAACAAGATTAAATGACAAGTGACTCTTAATGATGATCGTAGACTTC 3173  
Qy 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
Db 3174 TTTTAAGCAGTGGGTTGTTAAGTACATGACAAAGAACTGACAGCAACCAATGATCCACACC 3223  
Qy 401 TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlyAspPro 420  
Db 3234 TACCTGGACAAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTCTGAAAGACTCA 3293  
Qy 421 ProThrAlaValAlaAlaValAlaGluTTPheAspGlyLysAspPheGlnGlySerLysLeu 440  
Db 3294 CCACTGCCAAAGCCTCCGGGATGATGATGAGGAGAAAGATTTTCAAGAGGACCAAACTT 3353  
Qy 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
Db 3354 AAAGTCTCTTCTCTGGAAGAGGCTCCAGTAAACGATATGACAGGTGTATGACACCC 3413

QY	46	ArgIugIyAsgGIyMePProPProleuMArgGIyGIyPProGIyGIyPProGIyGIyPPro	480
Db	3414	CGTAgGCGCAAGGAGATCCACACCACTCTCGGAGGCTCAGAGAGGCCCAAGAAgTCTT	3477
QY	481	GIyGIyPProMeGIyArMeGIyGIyArGIyGIyAAsPArGIyGIyPhePProArG	500
Db	3474	GGGGAGCCCATGGGTCACATGGAGAGCCGTGGAGAGATAGAGAGGCCCTCCCTCCAGA	3533
QY	501	GIyPProArGIySeArGIyAAsPProSeArGIyGIyGIyAAsValGIhIbArGIaGIy	520
Db	3534	GGACCCCGAGGTTCCCGAGGGAACCTCTGGAGAGAGAAAGCTCAGACCAACCACTGGA	3593
QY	521	AsPTArgInCyBPProAsPProGIyCyAgGIyAAsGIyAAsnPhaPAlATrArGThrGIuCy	540
Db	3594	GACAGCACTGTCCCATTCGGGTTGTGGAAACCAAACTTGCTCGTAGAAACAAGAAC	3655
QY	541	AAsnGIuCybAlAProLyAsPProGIyPheLeuPProPProPhePProPProGIy	560
Db	3654	AAACAAGTTAAAGGCTCCAAAGCCTGAAGCCTTCCTCCGCAACCTTCCCAACCCCGGTT	3711
QY	561	GIyAsPArGIyArGIyGIyPProGIyGIyMeLArGIyGIyArGIyGIyGIyLeuMeLAsP	580
Db	3714	GGTATCATGCGACAGGAGTGCGCTTGCTGTGACATGTGGGAGGAAGGTGGCTCATGTGAT	3772
QY	581	ArgGIyGIyPProGIyGIyMePheArGIyGIyArGIyGIyAAsPArGIyGIyPheArG	600
Db	3774	CATGGTGGTCCCGGTGGATGTTCAAGAGTGCTGTGTAGACACAAATGGCTTCCGT	3833
QY	601	GIyGIyArGIyMeLAsPArGIyGIyPheGIyGIyGIyArGIyGIyGIyPProGIyGIy	620
Db	3834	GGTGGCTG-dGGATGTGAACGAGGTGGCTTTGGTGAAGGAAGAAGGTGGCCCTGGGGGG	3892
QY	621	PProPProGIyPProLeuMeGIuGIuMeGIyGIyArGIyGIyGIyArGIyGIyPProGIy	640
Db	3893	CCCCCGGA-CCTTTGATGTAAACAAATGGAGAGAAAGAGAGAGCTGAAGAACTTGA	3953
QY	641	LyMeLAsPArLyGIyGIyGIyArGIyGIyGIyArGIyGIyGIyArGIyGIyGIyPProGIy	656
Db	3952	AAACCGATAAAGCGAGCACTGTCAAGAGCCCAAGAGATCAACCCCTTAC	3999

```

/ RESULT 3
/ US-09-949-016-5043
/ Sequence 5043, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949, 016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5043
/ LENGTH: 1785
/ TYPE: DNA
/ ORGANISM: Human
/ US-09-949-016-5043

```

Alignment Scores:	
Pred. No.:	6,89e-162
Score:	3178, 50
Percent Similarity:	88,728
Best Local Similarity:	88,728
Query Match:	87,494
DB:	4
Length:	1765
Matches:	582
Conservative:	0
Mismatches:	1
Indels:	73
Gaps:	1

US-10-791-017A-2 (1-656) x US-09-949-016-5043 (1-1785)

Qy 1 MetLaserThrAspTyrSerSerThrTyrSerGlnIaaIaaIaGlnGlnGlyTyrSerAla 20

Db 34 ATGGCGTCACGGATTACAGTACCTATAGCCAAAGCTGACGCCACACGGGCTACAGTGCCT 93

Qy 21 TyrThrAlaGlnProThrGlnGlyTYrZalaglnThrThrGlnAlaTYrGlyGlnGlnSer 40

Db 94 TACACCGCCACGCCCATCTCAAGATATATCAACAGACACCCAGGCGATATGGCAACAAGC 153

Qy 41 TYrGlyThrTYrGlyGlnProThrAspValSerTYrThrGlnIaGlnThrThrAlaThr 60

Db 154 TATGGAACTATGACACAGCCACCTGATCTGACTATACCCAGGCTCAGACCACTGCACCC 213

Qy 61 TYrGlyGlnThrAlaTYrAlaAlaThrSerTYrGlyGlnProProThrGlyTYrThrThrPro 80

Db 214 TATGGGACAGCCGCTATGACACTTCTATATGGACGCCCTCCACGTGTATACTACTCCA 273

Qy 81 ThrAlaProGlnAlaTYrSerSerGlnProValGlnGlyTYrGlyThrGlyValaTYrAspThr 100

Db 274 ACTGCCCCCAGGCATACAGCCAGCTGTCCAGGGGTATGGACATGGTCTTATATGATACC 333

Qy 101 ThrThrAlaThrValThrThrThrGlnAlaSerTYrAlaIaGlnSerAlaTYrGlyTYr 120

Db 334 ACCACTGTACAGTACACCAACCAACCGAGCGCTCTCTATGACGACTCAGCTGCATATGCGACT 393

Qy 121 GlnProAlaTYrProAlaTYrGlyGlnGlnProAlaAlaAlaThrAlaProThrThrProGln 140

Db 394 CAGCTGTCTTATCCAGCCTATATGGGACGACGACGACGACGACCTATCAACAGACCGGACG 453

Qy 141 AspGlyAsnValProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTYrZaenGln 160

Db 454 GATGGAAACAAGCCCATCTGAGACTATGTAACCTCATATGACACAGGGGGTTTACAACCG 513

Qy 161 ProSerLeuGlyTYrGlyGlnSerAsnTYrSerTYrProGlnValProGlySerTYrPro 180

QY	201	ThSertYrAaPgiInSerSerTyrSerGlnGlnaantHrTyrGlyGlnProSerSerTyr	220
Db	634	ACTGTTATAGATCAGACGAGTTACTCTCAGCAAGAACCTTAGGGCAACCGACACTAT	633
QY	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr	240
Db	694	GGACAGCAGAGTAGCTATAGTCAACAAAGCAGCTATGGGCAGCAGCTCTCCACTAGTTAC	753
QY	241	ProProGlnInHrGlySerTyrTyrSerGlnAlaProSerGlnTyrTyrSerGlnInSerSerSer	260
Db	754	CCACCCCAACTGATCCTACAGCCAACTCCAACTCAATATAGCCACAGACAGACAC	813
QY	261	TyrGlyGlnGlnSerSerPheArgGlnAaspHisProSerSerMetGlyAlaTyrGlyGln	280
Db	814	TACGGGCGACGACAGA-----	828
QY	281	GluSerGlyGlyPheSerGlyProGlyGluAanaArgSerMetSerGlyProAaPaanaArg	300
Db	828	-----	828
QY	301	GlyAhrGlyAhrGlyGlyPheAaPaArgGlyGlyMetSerArgGlyGlyAhrGlyGlyGly	320
Db	828	-----	828
QY	321	ArgGlyGlyMetGlySerAlaGlyGlyAuaArgGlyGlyPheAaInLaProGlyGlyProMet	340
Db	829	-----CCCATG	834
QY	341	AspGluGlyProAaPleAaPleGlyProProValaAaPProAaSpGlyAaSpSerAaPaan	360
Db	835	GATGAAGACCAAGACTTGATCTTAGGCCCACTGTAGATCCAGATGMAAAGCTTGACAA	894

```

QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380
DB 895 AGTGCAATTATGATGACAAAGATTAAATGACAGTGTGCTGATGATGATCTGGCAGACTTC 954
QY 381 PheLeuGlnCysGlyValValIleMetAsnLysArgThrGlnIlePheMetIleHisIle 400
DB 955 TTTAAGCAGTGTGGGTGTTGTTAAGATGACAAAGAACTGGGCAACCAATGATCAATC 1014
QY 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAsnAlaThrValSerTyrGluAspPro 420
DB 1015 TACCTGACAAAGAAACAGAAAGCCAAAGGCATGCCACAGTCTCTATGAAGACCA 1074
QY 421 ProThrAlaIleValAlaValGlnTyrPheAspGlyLysAspPheGlnGlySerIleLeu 440
DB 1075 CCACATGCGCAAGGCTGCTGGTAATGTTGATGGAAAGATTTTCAAGGAGCAAACTT 1134
QY 441 LysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgGlyLysLeuProPro 460
DB 1135 AAGTCTCCTTCTGCTCGAAGAAAGCTCCATGAAACGTTATGGGGTGTCTCCACCC 1194
QY 461 ArgGluIleValArgLysMetProProProLeuArgGlyLysProGlyLysProGlyPro 480
DB 1195 CGTAGAGGCGAGAGCATGCCACCACTCCGTGAGAGTCCAGGAGGCCAGAGGCTCT 1254
QY 481 GlyGlyProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPheProProArg 500
DB 1255 GGGGAGACCCATGGGTGCTGATGGAGGCGCTGAGAGAAATAGAGAGGCTTCCCTCCAGA 1314
QY 501 GlyProArgLysSerArgLysAsnProSerGlyLysLysValGlnHisArgAlaGly 520
DB 1315 GAGACCCGGGGGTTCCGAGGGAACCTCTGAGAGAGAAACGTCAGCACAGAGTGA 1374
QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThrGlyCys 540
DB 1375 GACTGGCAGTGTCCCATCCGGGTGTGAAACCAAGAACTTCGCTGAGAAACAGAGTGC 1434
QY 541 AsnGlnCysLysValAlaProLysProGlnGlyLysPheLeuProProPheProProGly 560
DB 1435 AACCAAGTAAAGGCCCAAGGCTGAAAGGCTTCTCCGCGCACCTTCCGCCCCCGGGT 1494
QY 561 GlyAspArgLysArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeuMetAsp 580
DB 1495 GGTGATCTGTGGCAGAGGTGCTGCTGTGGCATGCGGGAAGAAAGGTGCTCATGAT 1554
QY 581 ArgGlyLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLysPheArg 600
DB 1555 CGTGTGTGTCCCGGTGAATGTTCAAGAGTGTGCGGTGTGGAACAGAGTGGCTTCGT 1614
QY 601 GlyLysArgLysMetAspArgGlyLysPheGlyLysLysArgGlyLysProGlyLys 620
DB 1615 GGTGGCGGGGCGATGACCGAGGTGCTTGTGTGAGAGAAACAGAGTGGCTTGGGGGG 1674
QY 621 ProProGlyProLeuMetGlnGlnMetGlyLysArgArgGlyLysArgGlyLysProGly 640
DB 1675 CCCCCGTGACCTTGTATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGACTTGA 1734
QY 641 LysMetAspLysGlyLysLysArgGlnGlnLysArgAspArgProTyr 656
DB 1735 AAAATGATAAAGCGCAGACCGTCAAGAGCGCAGAGATCGGCCCTTAC 1782

```

```

RESULT 4
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTUR, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044

Alignment Scores:
Pred. No.: 6,89e-162 Length: 1785
Score: 3178.50 Matches: 582
Percent Similarity: 88.72% Conservative: 0
Best Local Similarity: 88.72% Mismatches: 1
Query Match: 87.49% Indels: 73
DB: Gaps: 1

US-10-791-017A-2 (1-656) x US-09-949-016-5044 (1-1785)
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20
DB 34 ATGGCTCCAGGATTAACCTTACCTTATGACAGCTGACGAGGAGGCTTACAGTGTCT 93
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer 40
DB 94 TACACCGCCAGCCACTCAAGATATGACAGACCAACCGGCAATATGGGCAACAAAGC 153
QY 41 TyrGlyThrTyrGlyLysProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
DB 154 TATGGAACCTATGACAGCCACTGATGTCAATTAACCAAGGCTCAGACACTGCAACC 213
QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyLysProProThrGlyTyrThrThrPro 80
DB 214 TATGGGCAAGCCCTTATGACACTTCTTATGACAGCTTCCACTGTTTACTCTCA 273
QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100
DB 274 ACTGCCCCCAGGCAATACAGCCAGCGCTGTCAAGGGATATGCACTGTGTATGATACC 333
QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
DB 334 ACCACGCTACAGTCAACCAACCAAGGCTCTCTTATGACAGTCACTGATATGACACT 393
QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140
DB 394 CAGCTGCTTATCAGCTTATGGGCAAGCCAGCAAGCCACTGACCTACCAAGACCGCAG 453
QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyLysTyrAsnGln 160
DB 454 GATGAAACAGGCCACTGAGACTTATGACACTCATCTAGCAACAGGGGTTTCAACAG 513
QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
DB 514 CCCAGCTAGATATGACAGAGTAACTAACATTATCCCAAGGTACTCGGAGACTACCCC 573
QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
DB 574 ATGCAAGCATCACTGACCTCATCTTACCTTACCTTACCTTACCTTACCAAGCCG 633
QY 201 ThrSerTyrAspLysSerTyrSerGlnGlnAsnThrTyrGlyLysProSerSerTyr 220
DB 634 ACTAGTTATGATACAGCACTTACTCTGACCAAAACCTATGGCAACGACAGCACTAT 693
QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
DB 694 GACACAGAGTATGATGTCAACAAAGCATATATGGGCAAGCAAGCTCCACTAGTATAC 753
QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260

```



Db	754	CCACCCCAAACTGATCTCTACAGCCAAAGTCCAAATCAATATATAGCAACAGACGACGAC	813
QY	261	TYRGIYGLINGINSErSePheArGIuInAphIaPProSeSerSeMetGIYValTYrGIYGLN	280
Db	814	TACGGGCGACGAGAGA-----	828
QY	281	GIUSErGIYGLYPhSeSerGIYProGIYGLuAaNArGSeSerMeSeSerGIYProAaSPaNArG	300
Db	828	-----	828
QY	301	GIYArGIYArGIYGLYIYPhEAaPaRGIYGLYMeTSeArGIYGLYArGIYGLYGLY	320
Db	828	-----	828
QY	321	ArGIYGLYMeTGIYSeRlaGIYGLuArGIYGLYIYPhEAaMlyAPProGIYGLYProMeT	340
Db	829	-----CCCAATG	834
QY	341	ASPGIUGIYProASPLeuASPLeuGIYProPProValaAPProASPGLuASPSeRASPbN	360
Db	835	GATGAAGAACAAGATCTTGATCTAGACCCCACTGTAGATCCAGATGAAGAAGCTCTGAACAAC	894
QY	361	SeRaIaIeTYrValGINGLYLeuAaNApSeSeRValThIRleuASPASPLeuAaSPhe	380
Db	895	AGTCGAATTATATGTCMAAGATTAATATGACAGTGTGACTCTTGAATGATCTGGCAACATTC	954
QY	381	PhelYSGINCYeGIYValIYAlYSeMeCbaNlyBaRGTINGLYINProMeTIIeHISIIe	400
Db	955	TTTAAGCAGTGGGGGTGTTAAGATGAACAAGAGAACTGGGCAACCATGATCCACATC	1014
QY	401	TYrLeuASPlySGINThrGIYlyAPProIySGIYAAPaIaThValSeTYrGIuASPPro	420
Db	1015	TACCTGGACAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAACAACCA	1074
QY	421	ProThRaIaIYSAIaIaValGIuTrTPheASPGLYlyASPPhelGINGLYSeRlySeIu	440
Db	1075	CCCACTGCCAAGGCTGCCGTGAAATGGTTTATGGAAAAATTTTCAAGGAGCAAACTT	1133
QY	441	LYSValSeRleuAaIaRGLYlyAPProPProMeTaenSeMeTArGIYGLYleuProPro	460
Db	1135	AAAGTCTCCCTTGCTCGGAAGAAAGCTCCAATGAACAGATGCGGGGTGTCTGCCACCC	1194
QY	461	ArGIUGIYArGIYMeTProProProleuArGIYGLYProGIYGLYProGIYGLYPro	480
Db	1195	CGTGAAGGACAGAGGATGCCAACACACATCCCTGTGAGGTCCAGAGGCCACAGAGATCCT	1254
QY	481	GIYGLYProMeTGIYArGMetGIYGLYArGIYGLYIYAAPArGIYGLYIYPhEProProArG	500
Db	1255	GGGGAGCCCAATGGGTCCGATGGAGAGCCGTGAGAGATAGAGAGGCTTCCCTTCAAGA	1314
QY	501	GIYProArGIYSeArGIYASbProSeRGIYGLYIYASbValGINHISArGlaGIY	520
Db	1315	GGAGCCCGGGGGTTCGAGGGGAACCCCTCTGAGAGAGGAACGTCCAGACCGAGCTGGA	1374
QY	521	ASPTTPGINCYAPProASPProGIYCYbGIYASbGINASbPhelATTPArGTINGLYCYb	540
Db	1375	GACTGGCAGTGTCCCAATCCGGGGTGTGGAAACCAAGAACTTCGCTCGAGAGAAACAAGTGC	1434
QY	541	ASbGINCYbLYSAIaIaProLYAPProGIUGIYPhEluAPProPProPProPProProGIY	560
Db	1435	AACCAAGTGAAGGCCCCCAAGGCTTAAGGCTTCTCCGCAACCTTCCGGCCCCCGGGGT	1494
QY	561	GIYAAPArGIYArGIYGLYProGIYGLYMeTArGIYGLYArGIYGLYIYleuMeTASP	580
Db	1495	GGTATCGGTGGAGAGGTGGCCCTGTGTGCANCGGGGAGAGAAAGTGGCTCATGAT	1554
QY	581	ArGIYGLYIYProGIYGLYMeTPhEAaRGIYGLYArGIYGLYIYASPArGIYGLYIYPhEAaR	600
Db	1555	CGTGTGTGTCCCGAGTGAATGTTCAAGAGGTGCGCTGTGTGAGAGACAGAGTGGCTTCCGT	1614
QY	601	GIYGLYArArGIYMeTASPArGIYGLYIYPhElGIYGLYIYArArGIYGLYIYProGIYGLY	620
Db	1615	GGTGGCCGGGGATGACCGAAGTGTGCTTTGGTGTGAGAGAAACGAGGTGGCCCTGGGGGGG	1674

OY		621	TROPPOGILYPROLEMEIGLIMETGLIYLATGATGGLIYLARGLIYLIPROGLIY	640
DB		1675	CCCCGTGAACCTTTGATTGAAACAAGTGGAGAAGAAAGAGAGAGCGTGAGAGACTGGA	1734
OY		641	LysMetAspIlySGIylunhiatrgingluarigaapayprotyr	656
DB		1735	AAAATGATAAAGCGCAGCACCGTCAGAGCGCAAGATCGGCCCTAC	1782
<b>RESULT 5</b>				
	US-09-949-016-1954			
/	Sequence 1954, Application US/09949016			
/	Patent No. 6812339			
/	GENERAL INFORMATION:			
/	APPLICANT: VENTER, J. Craig et al.			
/	TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
/	FILE REFERENCE: CLO01037			
/	CURRENT APPLICATION NUMBER: US/09/949, 016			
/	PRIOR FILING DATE: 2000-04-14			
/	PRIOR APPLICATION NUMBER: 60/241,755			
/	PRIOR FILING DATE: 2000-10-20			
/	PRIOR APPLICATION NUMBER: 60/237,768			
/	PRIOR FILING DATE: 2000-10-03			
/	PRIOR APPLICATION NUMBER: 60/231,498			
/	PRIOR FILING DATE: 2000-09-08			
/	NUMBER OF SEQ ID NOS: 207012			
/	SOFTWARE: FastSeq for Windows Version 4.0			
/	SEQ ID NO 1954			
/	LENGTH: 1783			
/	TYPE: DNA			
/	ORGANISM: Human			
/	US-09-949-016-1954			
<b>Alignment Scores:</b>				
Fed. No.:	4,84e-143	Length:	1783	
Score:	2826.50	Matches:	533	
Percent Similarity:	83.08%	Conservative:	12	
Best Local Similarity:	81.25%	Mismatches:	38	
Query Match:	77.80%	Indels:	75	
DB:	4	Gaps:	1	
US-10-791-017A-2 (1-656) x US-09-949-016-1954 (1-1783)				
OY		1	MeAlasErThraApTySerThTyTSerGiNaLaAGInGIInGIyTrSeRaLa	20
DB		34	ATGCATCATGATTAAGATTAACCTTAAGCCAGAGTCAGCCAGCAGGCTACAGTCT	93
OY		21	TyrThrAlaGlnProThrGInGIyTYraGInThrThrGInaLYrGInGInSer	40
DB		94	TACAACACCAGCCACTCAAGATATGACAGAACCCAGGCAATATGGCAACAAAG	153
OY		41	TyrGlyThrTYrGlyGlnProThrAspValserTYrThrGInaAGInThrThraLatnr	60
DB		154	TATGGAACTTAGAGACAGCCCATGTAGTCAAGCTATACCAAGCTCAGCACTGCAATC	213
OY		61	TyrGlyGInThraLaTYrAlaThSerTYrGlyGInProToThGlyTYrThrThPro	80
DB		214	TATGGGAGACCGCTATGCAACTTCTTATGGACAGCCTCCACATGGTTATACTCCA	273
OY		81	ThraLaProGInaLaTYrSerGInProVaGInGIyTYrGlyThrGIyALaTYrAspThr	100
DB		274	ACTGCCCCCAGGCAATACAGCCAGGCTCTCCAGGGGTATGGACATGTCCTTATGATAC	333
OY		101	ThrThraLatnrValThThrThrThrGInaLaserTYrAlaAGInSerLaTYrGlyThr	120
DB		334	ACCACTGCTACAGTCACCAACCAAGGCTCTTCATATGACGCTCAATCTGCATATGGCACT	393
OY		121	GlnProaLaTYrProaLaTYrGlyGInGlnPProaLaLaThraLaProThraPProGln	140
DB		394	CAGCTGCTTATCCAGCTTATGGGAGAGCAAGCCAGTACGCACTTACMAAGCCGAG	453
OY		141	AspGlyenLySProThrGluThrSerGlnProGInSerSerThrdLyGlyTYraGIn	160

```

Db      454  AATGGAACAGCCCACTAGACCTAGTCAACCTGACACAGAGGGATTACACAG 513
      161  ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
      514  CCCAGCCTAGATATGAGACAGATGAACTGACGTTATCCACGATGACTGGAGACTATCC 573
      181  MetGlnProValThrAlaProProSerTyrProProThiSerTyrSerSerThrGlnPro 200
      574  ATCAGGCAATCCAGCACTCCATCTTACCTTCACAGCTATTCCTTACACAGCCA 633
      201  ThiSerTyrAspGlnSerSerTyrSerGlnAsnThrTyrGlyGlnProSerSerTyr 220
      634  ACTAGTTATATCAGACCACTTACTCTCAGACAGAACCTTATGGAGAAACGAGCACTAT 693
      221  GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThiSerTyr 240
      694  GACAGCAGAGTATGCTATGCTCAACAAAGCAGTATGGCAGCTGCCCTCCACTGATTAC 753
      241  ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260
      754  CCAACCCCAACTGATCTTACAGCCAAAGCTCCAAATGATATGCTTAAACAGAGCAGCAG 813
      261  TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
      814  TACGGCAGCAGAGA----- 828
      281  GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300
      828 ----- 828
      301  GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyArgGlyGlyGly 320
      828 ----- 828
      321  ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnArgProGlyGlyProMet 340
      829 ----- 828
      341  AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspAspAsn 360
      835  GATGAAGGACCAAGTCTTGATCTAGGCCCACTGATGATCAAGTGAAGACTGCAAC 894
      361  SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380
      895  AGTCAATTATATCAAAAGATTAAGATGACATGATGATCTAGATGATCTGATGACTTC 954
      381  PheAlaGlnCysGlyValValLeuMetAsnLeuAspArgThrGlyGlnProMetIleHisIle 400
      955  TTTAAGCAGTGTGGGTTGTTAGATGACAAAGAACTGACCAACCCATGATCCACACC 1014
      401  TyrLeuAspLeuGlnThrGlyGlyProGlyGlyAspAlaThrValSerTyrGlyAspPro 420
      1015  TACCTGGACAAAGAAACAAAGAAAGCCCAAGGTGATGTCACAGTGTCTCTGAAAGACTCA 1074
      421  ProThrAlaValAlaValAlaValGluTyrPheAspGlyValAspPheGlnGlySerLeu 440
      1075  CCTACTGCTCCAAAGCTGCTGGAATGTTGATGGGAAAGATTTTCAAGGAGCAAACTT 1134
      441  LysValSerLeuAlaValValValValProProMetAsnSerMetArgGlyGlyLeuProPro 460
      1135  AAGGTCTCTTCTCTGGAAGAGGCTCCAGTGAACGATGACAGGTTGATGACCACTCC 1194
      461  ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480
      1195  CGTAGAGGAGAGAGGAGGACCACTCTGCGAGGTCCAGAGGCCCAAGAGAGTCTT 1254
      481  GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500
      1255  GGGGAGCCCTGAGTCAATGAGGAGGCGGTGAGAGATGAGAGGCTCCCTCCCAAGA 1314
      501  GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyValAsnValGlnHisArgAlaGly 520
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1315  GACCCCAAGGTTCCCGAGGAAACCTCTGAGAGAGAAAGTCCACAGCAAGCTGGA 1374
      521  AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrPargThrGlyCys 540
      1375  GACAGCAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGGAAGACAGAGAGC 1434
      541  AsnGlnCysValAlaProLysProGlnGlyPheLeuProProProPheProProProGly 560
      1435  AACAGGTAAAGGCTCCAAAGCCTGAAGGCTTCTCCCGCACCTTCCACCCCGGGGT 1494
      561  GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580
      1495  GTGATCATGAGCAGAGGTGCTCTGTGTGATGTGGAGAGAGAGAGTGGCTCATGAT 1554
      581  ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyValAspArgGlyGlyPheArg 600
      1555  CATGTGTGTCCTCGGTGAATGTTCAAGAGGTGCTGTGTGAGAAACAGAAATGGCTTCGCT 1614
      601  GlyValArgGlyMetAspArgGlyGlyPheGlyGlyGlyValArgArgGlyGlyProGlyGly 620
      1615  GTTGGCTG-GGCATGACCGAGGTGCTTGTGTGAGAGAAAGACAAAGTGGCTTGGGGGG 1673
      621  ProProGlyProLeuMetGlnMetGlyGlyValArgArgGlyGlyValArgGlyGlyProGly 640
      1674  CCCCCGGA-CTTTGATGTAAACCAATGGAGAGAAAGAGAGAGACTGAAGACTTGA 1732
      641  LysMetAspLeuGlyGlyHisArgGlnGluArgArgAspArgProTyr 656
      1733  AAAACGATTAAGCGCAGCACTGTCAAGAGCGCAGAGATCAGCCCTTAC 1780

RESULT 6
US-08-437-027-18
; Sequence 18, Application US/08437027
; Patent No. 5670317
; GENERAL INFORMATION:
; APPLICANT: Landanyi, Marc
; TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
; TITLE OF INVENTION: SMALL ROUND CELL TUMOR
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,027
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 46416/JPW/CCA
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2413 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-437-027-18
Alignment..Scores:

```

Pred. No.: 1.35e-67 Length: 2412  
 Score: 1420.00 Matches: 265  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 39.09% Indels: 0  
 DB: 1 Gaps: 0

US-10-791-017a-2 (1-656) x US-08-437-027-18 (1-2412)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 Db 25 ATGGCGCTCCACGAGTTCAGTACCTATAGCCAGCTCAGGCGAGGCTCAGTGGCT 84  
 QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlnGlnSer 40  
 Db 85 TACAGCCGCCAGCCAGCTCAGATATGACAGACCCAGGCAATAGGCAACAGAGC 144  
 QY 41 TyrGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 Db 145 TATGGAACCTATGACAGCCAGCTGATGTCAGTATCCAGGCTCAGACCACTGCAACC 204  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrPro 80  
 Db 205 TATGGCAGACCGGCTATGCACTTCTATGAGACAGGCTCCAGCTGTTATCTACTCCA 264  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlnAlaTyrAspThr 100  
 Db 265 ACTGCCCCCGAGGATACAGCCAGCTGTCAGGAGATAGGAGTGGTCTATAGATACC 324  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 Db 325 ACCAGCTGTAAGTACACCAACCAAGCCCTCTATGAGTCACTGTCATATAGGACT 384  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspGln 140  
 Db 385 CAGCTGTTATCCAGCTATGAGGAGAGCAGCAGCAGCAGCTGCACTTACAGACCCAG 444  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
 Db 445 GATGAAACAGAGCCAGCTAGACCTAGTCACTTACCTATGACAGAGGGGTTACACAG 504  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 Db 505 CCCAGCCTAAGATATGACAGATATGACAGTATATCCCAAGTACCTGGAGTACCCC 564  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 565 ATGACGACAGTCACTGCACTCCATCTTACCTCTACAGCTATCTCTACACAGCCG 624  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 Db 625 ACTAGTATATGATCAGAGAGTATCTCTCAGCAAAACCTATAGGCAACCAAGAGCTAT 684  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 Db 685 GACAGAGAGATAGTATGTCACAAAGAGCTATAGGAGAGAGCTCCCACTAGTTAC 744  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 745 CCACCCCAAACTGATCTTACAGCAAGCTCAAGTCAATATACCAACAGAGAGAGC 804  
 QY 261 TyrGlyGlnGlnSer 265  
 Db 805 TACGGGAGAGAGAGT 819

# RESULT 7

US-09-919-039-322  
 : Sequence 322 Application US/09919039  
 : Patent No. 6727066  
 : GENERAL INFORMATION:  
 : APPLICANT: Kaber, Matthew R.  
 : TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
 : FILE REFERENCE: PA-0035 US

CURRENT APPLICATION NUMBER: US/09/919,039  
 : CURRENT FILING DATE: 2002-09-09  
 : PRIOR APPLICATION NUMBER: 60/222,113  
 : PRIOR FILING DATE: 2000-07-28  
 : NUMBER OF SEQ ID NOS: 401  
 : SOFTWARE: PERL Program  
 : SEQ ID NO 322  
 : LENGTH: 1939  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : NAME/KEY: misc feature  
 : OTHER INFORMATION: Incyte ID No. 6727066 478620.53  
 US-09-919-039-322

Alignment Scores:  
 Pred. No.: 7.51e-60 Length: 1939  
 Score: 1273.50 Matches: 301  
 Percent Similarity: 53.96% Conservative: 60  
 Best Local Similarity: 44.99% Mismatches: 153  
 Query Match: 35.05% Indels: 156  
 DB: 4 Gaps: 30

US-10-791-017a-2 (1-656) x US-09-919-039-322 (1-1939)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 Db 79 ATGGCGCTCCACGAGTTCAGTACCTATAGCCAGCTCAGGCGAGGCTCAGTGGCT 126  
 QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlnGlnSer 39  
 Db 127 TACCCACCAAGCCAGGAGGCTATCCAGAGAGAGTATGAGCTTACAGAGAGCAG 186  
 QY 40 SerTyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAla 59  
 Db 187 AGTAAAGTGGTATATGACAGTCCAGCAGAC-----ACTTCA 222  
 QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThr 78  
 Db 223 GGTATAGCCAGAGAGAGCTAT-----TCTTATAGCCAGAGCAACCAAGAGCTATGGA 279  
 QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAla 97  
 Db 280 ACTCAGTCAACTCCC-----CAGGAGATAGGCTTCACTGCGCGGC 318  
 QY 98 TyrAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAla 117  
 Db 319 TATGGC-----AGTACCAAGAGCTCCCAATGCTCT 348  
 QY 118 TyrGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137  
 Db 349 TACGGGAGAGAGCTCTCTCTCTGCTATGCGACAGCAGCT----- 393  
 QY 138 ArgProGlnAspLysAsnLysProThrGlnThrSerGlnProGlnSerSerThrGly 157  
 Db 394 -----CCAGAGAGCACTCCGGAAGT 414  
 QY 158 TyrAsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGly 177  
 Db 415 TACGGTATGAGTCT-----CAGAGAGAGCTATAGGAGAGCCCAAG-----AGTGGG 462  
 QY 178 SerTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSer 197  
 Db 463 AGCTACAGCAGAGCT-----AGCTATGAGCA 492  
 QY 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnPro 217  
 Db 493 -----CAGCAGCAAGCTATGACAGAGCAAGAGCTATATCCCT 534  
 QY 218 SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProPro 237  
 Db 535 CAGGAGTATGAGAGAGCAAGCAAGTACCAACAGAGAGTGTGTGAGAGGTGAGGT- 593

QY 238 ThSerTyRProProGlnThnGlySerTyRSerGlnAlaProSerGlnIntyRserGln 257  
 DB 594 GAGAGT-----GAGAGTACTATGCGCAATCAATCTTCATGAGTAGTGT 641  
 QY 258 SerSerSerTyRglnGlnSerSerPheArgGlnAphIaProSerSerMetGlyVal 277  
 DB 642 GGTGGAGGTGTGTGGTGTATGGCAATCAAGACAGAGTGTGGAGTGGCAGCGGTGC 701  
 QY 278 TyRglnGlnGlnSerGlyGlyPheSerGlyProGlyGlnAphSerMetSerGlyPro 297  
 DB 702 TATGGACAGACGACCGTGA-----GGCGCGGCGAGGGGTGGCAGGTGT--- 746  
 QY 298 AspAphArgGlyArgGlyArgGlyGlyPheAphArg-----GlyGlyMetSerArgGly 315  
 DB 747 GGGGGCGCGCGCGCGCGGTGTGTGTACACCGCACGTGTGTGTGAACCCAGA 806  
 QY 316 GlyArgGlyGlyGlyArgGlyGlyMetGlySerAlaGly-----GlyArgGlyGlyPhe 333  
 DB 807 GGTGTGGAGGTGTGGTGTGGAGCGAGGTGGCATGGCGCGAGTGAACCGTGTGTTC 866  
 QY 334 AsnLysProGlyGlyProMetAspGlnGlyProAphLysLysLysLysProValAsp 353  
 DB 867 AATTAATTTGT 911  
 QY 354 ProAphGlnAphSerAspAphSerAlaIleTyRValGlnGlyLysAphSerValThr 373  
 DB 912 GAAACGATATATCAACCAACCAACCAATCTTGTGTGTGTGTGTGTGTGTGTGTGT 971  
 QY 374 LeuAphAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 393  
 DB 972 ATTAGAGT 1031  
 QY 394 GlyLysProMetIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 413  
 DB 1032 GGACAGCCCATGATTAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1091  
 QY 414 ThrValSerTyRglnAphProProThraLysValAlaValGlnTTPheAspGlyLys 433  
 DB 1092 ACGGTCTTGT 1151  
 QY 434 AspPheGlnGlySerLysLysLysLysLysLysLysLysLysLysLysLysLys 453  
 DB 1152 GAATTCGCCGAATCTATCAAGGTCTCATTTGTGTGTGTGTGTGTGTGTGTGTGT 1208  
 QY 454 MetArgGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 473  
 DB 1209 ---CGGGGT 1250  
 QY 474 ProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMetGlyArgGlyGlyAsp 493  
 DB 1251 CCCATGGGCGGTGAGGCTATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1298  
 QY 494 ArgGlyGlyPheProProArgGlyProArgGlySerArgLysAphProSerSerGly 513  
 DB 1299 CGAGGAGGATTTCCCACTGAGGT-----GGTGGCGGT 1331  
 QY 514 AsnValGlnIleSarGlnGlyAspTyRglnCysProAphSerProGlyCysGlyAphGln 533  
 DB 1332 GAGAGACAGACGACGAGT 1391  
 QY 534 PheAlaLeuPheArgGlnGlyAsnGlnCysLysValaProLysProGlyGlyPheLysPro 553  
 DB 1392 TTCTCTTGGAGGATGATGACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1442  
 QY 554 ProProPheProProProGlyGlyAspArgGlyArgGlyGlyProGlyGlyMetArg--- 572  
 DB 1443 ---CCAGAG-----GGGGGACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1472  
 QY 573 GlyLysArgGlyGlyLysMetAspArgGlyGlyProGlyGlyMetPheArgGly----- 590  
 DB 1473 GGGGGTAACTACGCGGAGT 1532  
 QY 591 ---GlyArgGlyGlyAspArgGlyGlyPheArgGlyGlyArg---GlyMetAspArgGly 608

DB 1533 CGGGGCGCGCGCGGAGCCTTGAGGCTTCCAGGCGCGCGGTGTGTGGACAGAGT 1592  
 QY 609 GlyPheGlyGlyGlyArgGlyGlyProGlyGlyProProGlyProLysMetGlnGln 628  
 DB 1593 GCGCTT----- 1598  
 QY 629 MetGlyLysArgGlyGlyArgGlyGlyProGlyGlyMetAsp---LysGlyGlnHis 647  
 DB 1599 ---GGCTTGGCAAGTGTATTCAGAGGTGTGACAC 1631  
 QY 648 ArgGlnGlnArgArgAspArgProTyR 656  
 DB 1632 AGACAGATGCGACGAGAGGCGCTAT 1658  
 RESULT 8  
 US-09-949-016-16785  
 ; Sequence 16785, Application US/0949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: C1001307  
 ; CURRENT APPLICATION NUMBER: US/09/949, 016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 16785  
 ; LENGTH: 35784  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-16785  
 Alignment Scores:  
 Pred. No.: 1,27e-41 Length: 35784  
 Score: 958.00 Matches: 233  
 Percent Similarity: 29.75% Conservative: 5  
 Best Local Similarity: 29.12% Mismatches: 11  
 Query Match: 26.37% Indels: 554  
 DB: 4 Gaps: 6  
 US-10-791-017A-2 (1-656) x US-09-949-016-16785 (1-35784)  
 QY 402 LeuAspLysGlnThnGlyLysProLysGlyAspAlaThrValSerTyRglnAphProPro 421  
 DB 31405 TTGACGAAAGTTACAGGACCACTTAATGATCGGAGGTG-TCAATTGATGATGAT--- 31460  
 QY 422 ThrAlaLysAlaAlaValGlu-TTP----- 429  
 DB 31461 -----GTGGAGT 429  
 QY 430 -----PheAspGlyLysAspPheGlnGlySerLysLysValSerLeuAlaArgGly 447  
 DB 31506 TTGTCTTGTTCAGGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTGTGGA 31565  
 QY 447 LysProProMetAsnSerMetLysArgGlyLysLysLysLysLysLysLysLysLysLys 467  
 DB 31566 GAAGCCTCCAAATGAACAGTATGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 31625  
 QY 467 oProProLysArgGly----- 472  
 DB 31626 ACCACCACTCCGT 31685  
 QY 472 ----- 472  
 DB 31686 ACACTTCATACCTTGAGAAACTTGAATTATTAAGTGAAGAAATATTAATGTGTGTGT 31745



CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 16786  
LENGTH: 35784  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-16786

Alignment Scores:  
Pred. No.: 1,27e-41 Length: 35784  
Score: 958.00 Matches: 233  
Percent Similarity: 29.75% Conservation: 5  
Best Local Similarity: 29.12% Mismatches: 11  
Query Match: 26.37% Indels: 554  
Gaps: 6

US-10-791-017a-2 (1-656) x US-09-949-016-16786 (1-35784)

QY 402 LeuaplysgluthrlylybprolybalyaplaThrValSerTyrgluapProPro 421  
Db 31405 TTAGAGAAAGATTACAGCAGACCTAATGATCTGGAGTG-TCAATTGATGATGAT--- 31460  
QY 422 ThrAlaValAlaValGlu-TTP----- 429  
Db 31461 -----GTGAGTTGGTGAACAGGAGTACAGGGAGTAAATTGATGTTCTG 31505  
QY 430 -----PheaspGlylybapPheGlnGlySerTybLeuValSerLeuAlaArgly 447  
Db 31506 TTGCTCTGTTCCAGGAAAGATTTCAGAGGAGCAACTTAAATCTCCCTGCTCGAA 31565  
QY 447 blybProPomebAnserMetArglylyLeuProProArglylyArglyMetPr 467  
Db 31566 GAGGCTCCCATGAAAGTATGGGGGTGGTCTGCCACCCCGTAGGGCAGAGGCATGCC 31625  
QY 467 oProProLeuArgly----- 472  
Db 31626 ACCACCACTCCGTGAGGTACTTTTCTGAGCTCCTGATGCTTAAAGTTTTCAGT 31685  
QY 472 ----- 472  
Db 31686 ACACCTCATACCTTGAGAACTTGATTTAGAGTGAAGAAATATATAATTGTGTAG 31745  
QY 472 ----- 472  
Db 31746 AGTCAATACTAGACTATCGAGAGTAAAGATGTTTGTGGGAATTAAGAGAGAGA 31805  
QY 472 ----- 472  
Db 31806 AGAATGAGAGAGCTGAGAGCCACTGCTGCTCACTCCAGCTGCATTATTCAGCT 31865  
QY 472 ----- 472  
Db 31866 TTGGTGTGTCTGTATAGACATGCTTATTCCTTAAGAAATTTGGGAGTTCAAGCAGGTG 31925  
QY 472 ----- 472  
Db 31926 CAGTGGCTACGCTGTATATCCAGCACTTTGGAGGCTAGAGTGGGGATCACCTGAG 31985  
QY 472 ----- 472  
Db 31986 GTCAAGATTTAAGACAGCTGCGCAAGTGTGAATCTCCTCTACTAATAAATACA 32045  
QY 472 ----- 472  
Db 32046 AAAATTAGCCGGGTATGTGTGTGATGCTGTAAATCCAGCTACTCGGAGGCTGAGGCA 32105

QY 472 ----- 472  
Db 32106 GAGAAATTGTTGAATCTGGGGGGGTGAGAGTTGCACTGAGCAAAAGATCTGCCACTGCA 32165  
QY 472 ----- 472  
Db 32166 CTCAGCCTGGCAACAGTGTGAGACTCCCTCTCCAAAAAATTTGTTGGG 32225  
QY 472 ----- 472  
Db 32226 AGCTGTGTTCTGTAGAGACGTGAACAGCTCTCTCAGAGGAAGGGGCTGATGGCT 32285  
QY 472 ----- 472  
Db 32286 GAGCCACACGAAACACGGGACAGGTGAGGAAATGACAGCATATCTGTGGTT 32345  
QY 472 ----- 472  
Db 32346 TACTTATGATTTTATTTATTTCTTATAGCAATTTGGTCTACAGAAATGATTTGCTGT 32405  
QY 473 ----- 486  
Db 32406 TCTTGTGTTCTTGTGTAGTCCAGAGGCCCCAGAGGCTGTGGGGGACCATGGGTG 32465  
QY 486 gMetGlylyArglylylybapArglylylybProProArglylyProArglylySerAr 506  
Db 32466 CATGGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCG 32525  
QY 506 gGlybapProSerGlylylylybAnValGlnhSargAlaGlybAspTrpGlnCysProAs 526  
Db 32526 AGGAAACCTCTGAGAGAGAAACGTCCAGACCGAGCTGAGAGACTGGAGTGTCCAA 32585  
QY 526 nProGly----- 528  
Db 32586 TCCGTA-TGACTTGTCTTGGCAATTTGATACCTACGAGTGAAGCACCCCTCCCTCAC 32644  
QY 528 ----- 528  
Db 32645 CCCATCCCACTAGAGTGAATGCTCTGTAGAGAAAGAAATGATGACCTGTATGG 32704  
QY 528 ----- 528  
Db 32705 CTGGTTAGGACATGATCAGCAATTCATCTGACGCTTCAAGCCTTCTGAAGATTGATT 32764  
QY 528 ----- 528  
Db 32765 TGACCTGTCTGTGGGTCATATGCTGCTGAGGCTGTGCTTAAAGCATGGGTACATA 32824  
QY 528 ----- 528  
Db 32825 GATCTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 32884  
QY 529 ----- 533  
Db 32885 TACTTCAATTTGATTTCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 32944  
QY 533 nPheAlaTrpArgThrGluCysAnGlnCysbAlaProLybProGlylylyPheLeuPr 553  
Db 32945 CTTCGCTGAGAAACAGAGTGAACAGTGTAAAGCCCCCAAGCTTGAAGCTTCTCC 33004  
QY 553 oProProPheProPro----- 559  
Db 33005 GCCACCTTCCGCCCCGGGTAGTGCAGTTTCAATGATGTCCTCTGCTTCTGCT 33064  
QY 559 ----- 559  
Db 33065 GCTAAACCTCTTTCTTATTTGTGGCTGTGTAATGCAAGTTGCCCTCTGCTTAACAC 33124  
QY 559 ----- 559  
Db 33125 TTGAGTTGTGCTGCTCTCATTTCTAATTTGTAGCCCGATGCCGAGATTGAGTGAAGTG 33184

[illegible]

```

RESULT 11
US-09-621-976-13361
: Sequence 13361, Application US/09621976
: Patent No. 6639063
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, J. B.
: APPLICANT: Jobert, S.
: TITLE OF INVENTION: ESTs and Encoded Human Proteins.
: FILE REFERENCE: GENSET 054PR2
: CURRENT APPLICATION NUMBER: US/09/621,976
: CURRENT FILING DATE: 2000-07-21
: NUMBER OF SEQ ID NOS: 19335
: SOFTWARE: Patent.pm
: SEQ ID NO 13361
: LENGTH: 411
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-621-976-13361

```

Alignment Scores:	
Pred. No.:	5 55e-25
Score:	609.00
Percent Similarity:	99.15%
Best Local Similarity:	99.15%
Query Match:	16.76%
DB:	4
	Gaps: 0
	length: 411
	Matches: 116
	Conservative: 0
	Mismatches: 0
	Indels: 1
	Gaps: 0

US-10-791-01A-2 (1-656) x US-09-621-976-13361 (1-411)

OY 338 G1yPmEaApGluGlyPrCaSpLeuApLeuGlyProPrCaValApPrOApSpGluAp 357

Db 62 G3ACCCATGATGTAAGAGCAGCATCTTATCTTAGGCCCACTGTGATCCAGATGANAAG 121

OY 358 SerApAnSerAlaIleTyrValGlnGlyLeuAnApSerValThrLeuApApLeu 377

Db 122 TCTGCAACAAGTCATATTATGTCMAAGATTAAATGACAGGTGACTTAGTGTCTG 181

OY 378 AlaApPhePheLeuGlnCyGlyValValIlySMeAsnLysArgThrGlyGlnPmEa 397

Db 182 GAGACTCTCTTAAGCAAGTGTGGGTGTGTAAGATGACMAAGAACTGGCAACCATG 241







QY 257 GlnSerSerSerTyrglynglnInSerSerPhearglnAspHisProSerSerMetGly 276  
DB 61 CAGAGCAGCAGCTACCGGCGAGCAGATTCATTCGACAGCAGCCACCGATGAGCATGGGT 120  
QY 277 ValTyrglynglnInSerSerGlyPheSerGlyProGlynglnAlaGlnSerMetSerGly 296  
DB 121 GTTTATGGGCGAGGCTTGGAGATTTTCGAGACGAGAGAAACCGGAGCATGATGGC 180  
QY 297 ProAspAlaArgGlyArgGlyArgGlyPheAspArgGlyGlyMetSerArgGly 316  
DB 181 CCGATTAACCGGGCGAGGGGAGAGGGGGAATTGATGTGGAGGCATGACAGAGGTGGG 240  
QY 317 ArgGlynglnArgGlynglnMetGly 325  
DB 241 CCGGAGAGAGACCGCGGTGGAATGGGA 267  
RESULT 15  
US-09-919-039-323  
/ Sequence 323, Application US/09919039  
/ Patent No. 6727066  
/ GENERAL INFORMATION:  
/ APPLICANT: Kaseer, Matthew R.  
/ TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
/ FILE REFERENCE: PA-0035 US  
/ CURRENT APPLICATION NUMBER: US/09/919,039  
/ PRIOR FILING DATE: 2002-09-09  
/ PRIOR APPLICATION NUMBER: 60/222,113  
/ NUMBER OF SEQ ID NOS: 401  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 323  
/ LENGTH: 1684  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ OTHER INFORMATION: Incyte ID No. 6727066 1813444CB1  
US-09-919-039-323  
US-10-791-017a-2 (1-656) x US-09-919-039-323 (1-1684)  
Alignment Scores:  
Pred. No.: 1,21e-17 Length: 1684  
Score: 484.00 Matches: 177  
Percent Similarity: 40.68% Conservative: 61  
Best Local Similarity: 30.26% Mismatches: 173  
Query Match: 13.32% Indels: 176  
DB: 4 Gaps: 30  
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
DB 66 ATGGCCTCAACAGATATACC-----CAACAGCAACCCAAAGCTATGGGGC 113  
QY 21 TyThrAlaGlnProThrGlnGlyTyrAla----GlnThrThrGlnAlaTyrGlyngln 39  
DB 114 TACCCACCCAGCCCGGCGAGGCTATTCACAGCAGCAGTCCCTACGACAGCAG 173  
QY 40 SerTyrglynglnTyrGlynglnProThrAspValSerTyrThrGlnAlaGlnThrAla 59  
DB 174 AGTTACAGTGTATACCCAGTCCAGGAC-----ACTTCA 209  
QY 60 ThrTyrglynglnAlaTyrAlaThrTyrSerTyrglynglnProPro---ThrGlyTyrThr 78  
DB 210 GGCATATGGCCAGACACTAT---TCTTATATGGCCAGAGCCAGAACACAGGCTATGA 266  
QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrGlyAla 97  
DB 267 ACTCAGCACTCCC-----CAGGATATGGCTCGACTGGGGC 305  
QY 98 TyAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAla 117  
DB 306 TATGGC-----AGTACCGAGAGCTCCCATCTCT 335

QY 118 TyrglyThrGlnProAlaTyrProAlaTyrGlynglnProAlaAlaThrAlaProThr 137  
DB 336 TACGGGAGAGATCTCCCTATCTGCTATGGCCAGACAGCAGCT----- 380  
QY 138 ArgProGlnAspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGly 157  
DB 381 -----CCAGCAGACACTGGGAACT 401  
QY 158 TyraGlnProSerLeuGlyTyrglynglnSerAsnTyrSerTyrProGlnValProGly 177  
DB 402 TAGGATAGCAGTCT-----CAGAGCAGAGCTATGGGAGCCCGA---AGTGG 449  
QY 178 SerTyProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSer 197  
DB 450 AGCTACAGCCAGCAGCT-----ACCTATGGTGA 479  
QY 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlynglnPro 217  
DB 480 -----CAGCAGAAAGCTATGACAGCAGCAAGCTATTAATCCCT 521  
QY 218 SerSerTyrglynglnInSerSerTyrglynglnInSer----- 230  
DB 522 CAGGCTATGACAGCAGAAACAGTACACAGAGCAGTGTGGAGGTGAGGTGCA 581  
QY 231 -----SerTyrglynglnInProPro-----ThrSerTyrProProGlnThrGly 245  
DB 582 GGTGAGAGTAACTATGGCCAAAGATCAATCTCCATGATGTGGTGGCAGTGTGGC 641  
QY 246 SerTySerGlnAlaProSerGlnTyrSerGlnGlnSerSerTyrglynglnInser 265  
DB 642 GGTATGGCANTACAGACAGAGTGTGAGTGGCGCGGCTATGACAGCAGAC 701  
QY 266 SerPheArgGlnAspHisProSerSerMetGlyValTyrglynglnInSerGlyPhe 285  
DB 702 CGTGAACGCCGCGGCGAGGGGTGCAATGCTGGC-----GGCGCGGGCGC 746  
QY 286 SerGlyProGlynglnAlaAsnArgSerMetSerGlyProAspAlaArgGlyArgGly 305  
DB 747 GCGGTGTGTGTACACCGCAGCAGTGTGTGTATGAAACCCAGAGTCTGTGA----- 800  
QY 306 GlyPheAspArgGlynglnMetSerArgGlyArgGlyGlynglnArgGlynglnMetGly 325  
DB 801 -----GGTGGCGG---GAGGCGAGAGTGGCAGTGGC 830  
QY 326 SerAlaGlynglnArgGlyPheAsnLysProGlynglnProMetAspGly----- 342  
DB 831 GGAAGT---GACCGTGTGCTTCAATAATTGTGTGTCAAGAAAGATATCTT 887  
QY 343 -----GlyProAspLeuAspLeuGlyProProValAspProAspGlynglnAspAsp 360  
DB 888 CATACATCACACACCTGAAGAGAGATGTGCTTTCCAGACTGATCAACTGAGATG 947  
QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
DB 948 GACGTGAGTCAATGCTTCTTCTTGGGACACTGTCCAGCTGGAGCTGGAAGCTGG 1067  
QY 381 Phe-LysGlnGlyValValAlaLysMetAsnLysAspGlnGlynglnProMetIleHis 400  
DB 1008 TATGAGAGCTGCAAGAGAGCTGTCTTCAGATGAAGAAATGGGGGTACTATGTTTCACT 1067  
QY 400 eTyLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrglynglnPro 420  
DB 1068 C---CTGGAATGAAAGAGAAAGATCAAAATCTTC--ACCACACTT-----GACCC 1114  
QY 420 oProThrAlaLysAlaValAlaGlnThrPheAspGlyLysAspPheGlnGlySerLys 440  
DB 1115 T-----GCTTCTGTGCTGTGCTGACTGAGAGAGACCAAGCAGCA----- 1157  
QY 440 uTyValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlynglnProPro 460  
DB 1158 -GAGGTCAACAGACCTCCACAGACCTCA--CTCTCCAGATTCCATCAGAGCTCTCT-- 1213

```
QY 460 oArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGly-- 479
Db 1214 -----GGCTCAGAGAGAGAGAGAGAG 1236
QY 480 -----ProGlyGlyProMe 484
Db 1237 ACCAAGGAGAACCGAAGAACGAGAGTGTCAATCCCGAGCCCGGGCTGGAAGC 1296
QY 484 tGlyArgMetGlyGlyArgGlyGlyAsp----- 493
Db 1297 AGCGCATGAAGAGAGAACGAGAGATGAAGAGAAAGTGACACAGCTAGCTGAAGAGA 1356
QY 494 -----ArgGlyGlyPheProProArgGlyProArgGlySerArgGlyAsn-- 508
Db 1357 ATGAACGGCTCAAGCAGGAATCGAGCGCT-----GACCAAGGAGATGAGCGCACTC 1410
QY 509 -----ProSerGlyGly-----GlyAsnValGlnHisAr 518
Db 1411 GCCGAGCTGTGATTGACCGAATGTGATCTGCACCAAGCATGAACAATTGGAGCATCA 1470
QY 518 gAlaGlyAspTyr 522
Db 1471 GTCCCCCACTTGG 1483
```

Search completed: February 21, 2005, 04:22:17  
Job time : 451.042 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 21, 2005, 02:07:46 ; Search time 3629.78 Seconds

(without alignments)  
1068.188 Million cell updates/sec

Title: US-10-791-017a-2

Perfect score: 3633

Sequence: 1 MASTDSTYSQAAAGQGYSA.....GGPGKDKGHRDRDRDPY 656

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Dgapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODE=frame+ p2n.model -DEV=xlp
-O/cgnt1/USPTO.spool.p/US10791017/runat.17022005.125809.22196/app.query.fasta.1.1358
-DB=Published Applications NA -OPT=fastp -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10791017 @CGN 1 1 879 @runat.17022005.125809.22196
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LENGTH=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-FAPOP=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7
```

Database: Published Applications\_NA:\*

```
1: /cgnt2_6/prodata/1/pub/pna/US07_PUBCOMB.seq:*
2: /cgnt2_6/prodata/1/pub/pna/PCT_NEW_PUB.seq:*
3: /cgnt2_6/prodata/1/pub/pna/US06_NEW_PUB.seq:*
4: /cgnt2_6/prodata/1/pub/pna/US06_PUBCOMB.seq:*
5: /cgnt2_6/prodata/1/pub/pna/US07_NEW_PUB.seq:*
6: /cgnt2_6/prodata/1/pub/pna/PCTUS_PUBCOMB.seq:*
7: /cgnt2_6/prodata/1/pub/pna/US08_NEW_PUB.seq:*
8: /cgnt2_6/prodata/1/pub/pna/US08_PUBCOMB.seq:*
9: /cgnt2_6/prodata/1/pub/pna/US09_PUBCOMB.seq:*
10: /cgnt2_6/prodata/1/pub/pna/US09_PUBCOMB.seq:*
11: /cgnt2_6/prodata/1/pub/pna/US09C_PUBCOMB.seq:*
12: /cgnt2_6/prodata/1/pub/pna/US09C_NEW_PUB.seq:*
13: /cgnt2_6/prodata/1/pub/pna/US10_PUBCOMB.seq:*
14: /cgnt2_6/prodata/1/pub/pna/US10_PUBCOMB.seq:*
15: /cgnt2_6/prodata/1/pub/pna/US10C_PUBCOMB.seq:*
16: /cgnt2_6/prodata/1/pub/pna/US10C_PUBCOMB.seq:*
17: /cgnt2_6/prodata/1/pub/pna/US10E_PUBCOMB.seq:*
18: /cgnt2_6/prodata/1/pub/pna/US10F_PUBCOMB.seq:*
19: /cgnt2_6/prodata/1/pub/pna/US10F_NEW_PUB.seq:*
20: /cgnt2_6/prodata/1/pub/pna/US11_NEW_PUB.seq:*
21: /cgnt2_6/prodata/1/pub/pna/US60_NEW_PUB.seq:*
22: /cgnt2_6/prodata/1/pub/pna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2390	9 US-09-880-107-3769	Sequence 3769, Ap
2	3633	100.0	2390	10 US-09-960-706-1081	Sequence 1081, Ap
3	3633	100.0	2390	10 US-09-873-319-717	Sequence 717, App
4	3633	100.0	2390	18 US-10-791-017a-1	Sequence 49, Appl
5	3605	99.2	2166	9 US-09-822-830A-49	Sequence 49, Appl
6	3483	95.9	2273	9 US-09-822-830A-410	Sequence 410, App
7	3277	90.2	1988	17 US-10-094-749-1074	Sequence 1074, Ap
8	1282	35.3	1824	17 US-10-439-703-58	Sequence 58, Appl
9	1281.5	35.3	1832	18 US-10-755-889-649	Sequence 649, App
10	1273.5	35.1	1939	10 US-09-919-039-322	Sequence 322, App
11	1219	33.6	763	18 US-10-425-115-182496	Sequence 182496,
12	1181	32.5	2299	14 US-10-198-846-9847	Sequence 9847, Ap
13	961	26.5	550	10 US-09-918-995-27690	Sequence 27690, A
14	868.5	23.9	568	10 US-09-918-995-9557	Sequence 9557, Ap
15	839	23.1	540	9 US-09-864-761-8001	Sequence 8001, Ap
16	774.5	21.3	505	9 US-09-864-761-7766	Sequence 7766, Ap
17	718.5	19.8	345	9 US-09-925-301-669	Sequence 669, Ap
18	684	18.8	401	10 US-09-918-985-3868	Sequence 3868, Ap
19	673	18.5	399	9 US-09-864-761-24705	Sequence 24705, A
20	631	17.4	417	9 US-09-960-352-10259	Sequence 10259, A
21	593	16.3	361	9 US-09-960-352-3851	Sequence 3851, Ap
22	496	13.7	521	9 US-09-933-797-481	Sequence 481, App
23	493	13.6	503	9 US-09-864-761-8728	Sequence 8728, Ap
24	493	13.6	503	16 US-10-029-386-4463	Sequence 4463, Ap
25	486	13.4	262	16 US-10-029-386-18163	Sequence 18163, A
26	486	13.4	1682	17 US-10-172-118-1872	Sequence 1872, Ap
27	486	13.4	1682	17 US-10-342-887-1872	Sequence 1872, Ap
28	484	13.3	1684	10 US-09-919-039-323	Sequence 323, App
29	475	13.1	253	9 US-09-864-761-25438	Sequence 25438, A
30	470.5	13.0	3061	15 US-10-101-510-602	Sequence 602, App
31	460	12.7	457	18 US-10-469-285-488	Sequence 488, App
32	423	11.6	1401	17 US-10-437-963-89443	Sequence 89443, A
33	420	11.6	1397	17 US-10-425-114-31980	Sequence 31980, A
34	420	11.6	1397	18 US-10-425-115-42134	Sequence 42134, A
35	414	11.4	483	10 US-09-918-995-2865	Sequence 2865, App
36	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
37	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
38	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
39	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
40	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
41	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
42	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
43	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
44	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App
45	400.5	11.0	5460	15 US-10-101-510-602	Sequence 602, App

## ALIGNMENTS

```
RESULT 1
US-09-880-107-3769
; Sequence 3769, Application US/09880107
; Patent No. US2002042981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3769
```

```

?   LENGTH: 2390
?   TYPE: DNA
?   ORGANISM: Homo sapiens
?   FEATURE:
?   OTHER INFORMATION: Genbank Accession No. US20020142281A1 X66959
US-09-880-107-3769

```

Alignment Scores:	
Pred. No.:	3,666-253
Score:	3633.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	9
Length:	2360
Matches:	656
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-791-017A-2 (1-656) x US-09-880-107-3769 (1-2390)

Qy	1	MetAlSeThApbYrSerThYrSerGlnAlaAlaGlnGlnGlnYrSerAla	20
Dp	44	ATGGCTCCACGGATTACAGTACCTTATGCCAGCTGACGGCGACAGGGCTTACAGTGCT	1030
Qy	21	TYrThAlaGlnProThrGlnGlyTYrAlaGlnThrGlnAlaTYrGlyGlnInSer	40
Dp	104	TACACGGCCACGGCCCTACAGAAATGACAGCAACCCAGGAAATATGGCAACAAGC	1630
Qy	41	TYrGlyThrTYrGlyGlnProThrAlaSerTYrThrGlnAlaGlnThrThrAlaThr	60
Dp	164	TATGGAACTTATGAGACGCCCTGATGTGACCTTATACCAGGGCTCAGACCACTGCACACC	2230
Qy	61	TYrGlyGlnThrAlaTYrAlaThrSerTYrGlyGlnProProThrGlyTYrThrThrPro	80
Dp	224	TATGGCAGACGGCCCTATGCACTTATATGACAGCTCCCACTGGTTATATCACTCCA	2830
Qy	81	ThrAlaProGlnAlaTYrSerGlnProValGlnGlyTYrGlyThrGlyAlaTYrAspThr	100
Dp	284	ACTGCCCCCAGGCACATACAGCAGGCTGTCCAGGGGTATGGCACTGGCTTATATGATACC	3430
Qy	101	ThrThrAlaThrValThrThrThrGlnAlaSerTYrAlaAlaGlnSerAlaTYrGlyThr	120
Dp	344	ACCACGTCTACAGTACCAACCAAGGCTCTTATGACCTCAAGTCTGCATATATGGCACT	4030
Qy	121	GlnProAlaTYrProAlaTYrGlyGlnInProAlaAlaThrAlaProThrArgProGln	140
Dp	404	CAGCTGCTTATTCAGCTTATGAGGACAGCAGCAGCACTGCACCTTACAAAGACGGCAG	4630
Qy	141	AspGlyAsnLysProThrGlnThrSerGlnProGlnInSerSerThrGlyGlyTYrAsnGln	160
Dp	464	GATGGAAACAGGCCCACTGAGCATAGTCAACCTCATATACACAGGGGGTTTAAACACAG	5230
Qy	161	ProSerLeuGlyTYrGlyGlnSerAsnTYrSerTYrProGlnValProGlySerTYrPro	180
Dp	524	CCAGCCTTATGATATGAGACAGATACATCAAGTTATCCCAAGTATCTGGGAGCTACCCC	5830
Qy	181	MetGlnProValThrAlaProProSerTYrProProThrSerTYrSerSerThrGlnPro	200
Dp	584	ATGCACCCAGTACATGCACCTCAATCTCAACCTCTTACCACTATTCCTTACACACACCG	6430
Qy	201	ThrSerTYrAspGlnSerSerTYrSerGlnGlnAsnThrTYrGlyGlnProSerSerTYr	220
Dp	644	ACTAGTTATGATCAGACAGTAACTCTCAGCAAAACCTATATGGGCAACCAAGACAGCTAT	7030
Qy	221	GlyGlnGlnInSerSerTYrGlyGlnGlnInSerSerTYrGlyGlnGlnInProProThrSerTYr	240
Dp	704	GAGACGCAAGTATGCTATGTGTCAAAAGACGCTATGGGACAGACCTCCCACTATGTTAC	7630
Qy	241	ProProGlnThrGlySerTYrSerGlnAlaProSerGlnTYrSerGlnGlnInSerSerSer	260
Dp	764	CCACCCCAACTGATCTCTACAGCCAGCTCCAGTCAATATACCAACAGAGACAGCAGC	8230
Qy	261	TYrGlyGlnGlnInSerSerPheArgGlnAspArgProSerSerMetGlyValTYrGlyGln	280
Dp	824	TACGGGACACAGATTCAATTCGACAGGACCAACCCCAATACATATAGGGGTTTATATGGGACG	8830

QY	281	GIUSeRgIyglYpHeSeRgIyProglYglUaSnRgSeRMeSeRgIyProaSPaSnRg	300
Db	884	GAGTCGTGAAGATTTTCCGGACCAAGAGAAACCGACATGAGTGGCCCTGATAAACGG	943
QY	301	GIyARgIyARgIyglYpHeaSPARgIyglYMeSeRaRgIyglYARgIyglYglY	320
Db	944	GGCAGGGGAAAGGGGGGATTTGATCGTGAAGCATGACAAAGGTGGCGGGAGAGAGA	1003
QY	321	ARgIyglYMeTglYSeRaIAGIyglUaRgIyglYpHeaSnlySProglYglYProMeT	340
Db	1004	CGCGGTGGAAAGGGCAGCGCTGAGAGAGAGAGTGGCTTCAATTAAGCCTGTGGACCCATG	1063
QY	341	ASpglUgIyProaSPleuASpLeuGlYProProVaIaSPProaSPglYuaSPSeRaSPaSn	360
Db	1064	GATAAAGAACCAAGATCTTGATCTTAAGCCCTCTTGAGTCCAAGTAAGAACTCTGAACAAC	1122
QY	361	SeRaIaIeTyVaIAGInglYleuaSnASpSeRaIThIleuaASpASpLeuaIaASpHe	380
Db	1124	AGTCAATTTATGTAACAAGATTAAATGACGTGACTCTAGATGATCTGGCAGACTTC	1188
QY	381	PhelYegInCYgSIyVaIyVaIyMeCaSnlySaRgThRgIyGlInProMeTIlEhSIle	400
Db	1184	TTTAAGCAGTGGGGGTGTTAATATGAACAAGAAACTGGGCAAACCATGATCAATC	1243
QY	401	TyRleuASpIySgIyThThRgIyLySPProLySgIyASpAlaThRValSeTyGlUaSPPro	420
Db	1244	TACCTGGCAAGAAACAGAAACCCAAAGCCATGCAAGTGTCTTAAGAAGACCA	1303
QY	421	ProThRaIaLySaIaAlaVaIGIyThRPhaSPglYySaEPheGInglYSeRlySeu	440
Db	1304	CCCACTGCAAGGCTGCGGTGAATGGTTTCAAGAAAGTTTCAAGSAGCAAACTT	1363
QY	441	LySaISeRleuaIaRgLySaIySPProMeRaSnSeRaRgIyglYleuProPro	460
Db	1364	AAAGTCTCCCTGCTCGGAAGAAACCTTCAATGAACGTAAGCGGGGGTCTGCAACC	1422
QY	461	ARgIyglYARgIyMeSPProProProleuaRgIyglYProglYglYProglYglYPro	480
Db	1424	CGTAAGGCAAGGCATGACCAACACTCGTGAAGTCCAGAGGCGCCAGAGGTCTT	1483
QY	481	GIyglYProMeTGIyARgMeTGIyglYARgIyglYASpARgIyglYPhaSPProRar	500
Db	1484	GGGGGACCCATGGGTGCGATGGAGGCGGTGAGAGATGAAGAGGCTTCCCTCCAA	1544
QY	501	GIySPRoARgIySeRaRgIyASnProSeRgIyglYglYASnVaIglInIaRgAlaGlY	520
Db	1544	GGAACCCGGGGTTCGCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGCTGA	1603
QY	521	ASpThRgInCYSPRoASnProRgIyCYAGIyASnIaSnPhaIaATPhaRgThGlUyS	540
Db	1604	GACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCGCTGGAGAAACAGAGTC	1663
QY	541	ASnGlInCYSaIySaIaProLySPProGlInGlYpHeuSPProSPProPheSPProProglY	560
Db	1664	AACCAAGTGAAGGCCCAAGACCTGAAGGCTTCTCCGCAACCTTTCGCCCCCGGGT	1722
QY	561	GIyASpARgIyARgIyglYProglYglYMeARgIyglYARgIyglYleuMeCaSP	580
Db	1724	GGTATGTGGCAAGGTGGCCCTGGTGGCATGGGGGAGGAAGAAGTGGCTCATGGAT	1783
QY	581	ARgIyglYProglYglYMeCaPheARgIyglYARgIyglYASpARgIyglYpHeaRg	600
Db	1784	CGTGTGTGTCGGGTGGAATGTTCAAGGTGGCGGTGTGAGACAGAGGTGGCTTCGGT	1843
QY	601	GIyglYARgIyMeCaSPARgIyglYpHeGlYglYglYARgARgIyglYProglYglY	620
Db	1844	GGTGGCGGGGCATGACCAAGAGTGGCTTTGTGTGAAGAAACAAAGGTGGCCCTGGGGGG	1903
QY	621	ProSPRoGlYSPRoIeMeSeTglInMeTglYglYARgARgIyglYARgIyglYProglY	640
Db	1904	CCCCCTGGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGACGTGGAGACCTGGA	1966
QY	641	LyMeCaSPlySgIyIuInIaRgIyglInIuATgRaSPaRgProTyT	656

```

Db      1964 AAAATGATTAAGCGACGACCGTCAGAGCGCAGAGATGCGCCCTAC 2011
RESULT 2
US-09-960-706-1081
/ Sequence 1081, Application US/09960706
/ Publication No. US20030134280A1
/ GENERAL INFORMATION:
/ APPLICANT: Munger, William E.
/ TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
/ FILE REFERENCE: 44921-5029-01US
/ CURRENT APPLICATION NUMBER: US/09/960,706
/ PRIOR FILING DATE: 2001-09-24
/ PRIOR APPLICATION NUMBER: 60/223,323
/ PRIOR FILING DATE: 2000-08-07
/ PRIOR APPLICATION NUMBER: 09/873,319
/ PRIOR FILING DATE: 2001-06-05
/ NUMBER OF SEQ ID NOS: 1124
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 1081
/ LENGTH: 2390
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
US-09-960-706-1081
Alignment Scores:
Pred. No.: 3,666-253 Length: 2390
Score: 3633.00 Matches: 656
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-791-017a-2 (1-656) x US-09-960-706-1081 (1-2390)
QY      1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrSerAla 20
Db      44 ATGGCGTCCAGCGATTCAGTACCTTATGACCAAGCTCAGCGCAGCGGCTACAGTGTCT 103
QY      21 TyrThrAlaGlnProThrGlnGlnTyrAlaGlnThrGlnAlaTyrGlnGlnGlnSer 40
Db      104 TACACCGCCCGACGCTCAAGATATGACAGACCCAGCGCATATGCGCAACAAAGC 163
QY      41 TyrGlyThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60
Db      164 TATGGAACCTATGAGACGCCCACTGATGTCACTATCCAGGCTCAGACCTGCAACC 223
QY      61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProThrGlyTyrThrThrPro 80
Db      224 TATGGCAGACCGCCCTATGCACTTCTTATGACAGGCTCCCACTGTTATTACTACCA 283
QY      81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyAlaTyrAspThr 100
Db      284 ACTGCCCGCCAGGATACAGCCAGCTGTCTCAGGCGATGAGCACTGTGTCTTATGATAAC 343
QY      101 ThrThrAlaThrValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
Db      344 ACCACGTCTACATGACCAACCAACCGGCTCTCTATGACGCTCATGTGCAATGGCACT 403
QY      121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln 140
Db      404 CACCGCTCTATCCAGCTTATGGCAGACCCAGCAGCACTGCACTTACAAAGCCGACG 463
QY      141 AspGlyAsnTyrAspProThrGlnTyrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160
Db      464 GATGGAAGAACAGCCCACTGAGATCTAGTCACTCAATCTAGCAACAGGCGGTTTCAACACAG 523
QY      161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
Db      524 CCACGCTTAGATATGAGACAGATTACTTAAGTTATCCCAAGTAACTCTGGAGACTACCCC 583

```

```

QY      181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db      584 ATGACAGCACTGACCTGACCTTCATCTTACCTTCAACGAGTATTCCTTACACACCGC 643
QY      201 ThrSerTyrAspGlnSerSerTyrSerGlnAlaAsnThrTyrGlnGlnProSerSerTyr 220
Db      644 ACTAGTTATATATAGACAGCTTACTCTCAGACAAACCTTATGGGCAACCGACAGCTAT 703
QY      221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db      704 GACAGCAGAGTATGATATGCTTACAAACAAAGCAGCTATGGCAGCAGCTCCCACTAGTTAC 763
QY      241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260
Db      764 CCAACCCCAACTGATCTTACACCAAGCTCCAAAGTCMAATATGACCAACAGACAGCAGC 823
QY      261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
Db      824 TACGGCAGCAGAGCTTCACTTCCAGACAGACCCCACTGATGCTGAGTGTATTTATGGCAG 883
QY      281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300
Db      884 GAGTCTGAGAGATTTTCCCGGACCGAGAGAGAACCGAGCATGATGGCCCTGATTAACCGG 943
QY      301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGlyGly 320
Db      944 GGCAGGGGAGAGAGGGGATTTGATCTGAGGATGAGCAGAGATGGGCGGGAGAGAGGA 1003
QY      321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnTyrProGlyGlyProMet 340
Db      1004 CGCGGTGAGATGGGACGCGCTGAGAGACGAGGTGGCTTCAATTAAGCTGTGTGAACCATG 1063
QY      341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn 360
Db      1064 GATGAAGACCAAGATCTTGTATAGGCGCTCTGTATGATCAAGATGAAGACTCTGACAAAC 1123
QY      361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380
Db      1124 AGTCAATTTATATTAACAAGATTAATGAACAGTGTACTCTTAATGATGTGGCAGACTTC 1183
QY      381 PheLeuGlnCysGlyValValLeuMetAsnLeuValThrGlyGlnProMetIleHisIle 400
Db      1184 TTTAAGCATGTGTGGGTGTTTAAAGATGAACAAAGAACTGGGGAACCAATGATTCACATC 1243
QY      401 TyrLeuAspGlyGlnThrGlyGlyProGlyGlyAspAlaThrValSerTyrGlyAspPro 420
Db      1244 TACCTGACAGAGAAACAGGAAAGCCCAAGGAGTCCCACTGTCTTATGAAGACCA 1303
QY      421 ProThrAlaValAlaAlaValGluTTPPheAspGlyValAspPheGlnGlySerLeu 440
Db      1304 CCCACTGCCAAGGCTCCCGTGAATGTTGATGGGAAAGATTTTCAAGGAGCAACAACTT 1363
QY      441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460
Db      1364 AAAGTCTCCCTGCTCGAAGAAAGCTTCAATGAACATATGAGGAGGTGTCTGCCACC 1423
QY      461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro 480
Db      1424 CGTGAGGAGCAGAGGATGCAACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGTCT 1483
QY      481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500
Db      1484 GGGGAGACCATGGGTGCAATGGAGGCGCGTGGAGAGATGAAGAGAGGCTTCCCTCCAGA 1543
QY      501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520
Db      1544 GAAACCCCGGGGTTCCTCGAGAGAAACCTCTCGAGAGAGAAACCTCCAGCAGAGCTGGA 1603
QY      521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaThrArgThrGluCys 540
Db      1604 GACTGCAAGTGTCCCAATCCGGTGTGGAAACCAACAATTCCTGCGTGGAGAACAGAGTGG 1663

```



Qy 541 AsnGlnCysIysAlaProLysProGluGlyPheLeuProProPheProProGly 560  
 Db 1664 AACCAAGTAAAGGCCCAAGAGCCCTGAAAGCTTCTCCCGCACCCTTCCGCCCCGGGT 1723  
 Qy 561 G1YAAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 Db 1724 GGTATGCTGGCAGAGGTGGCCCTGTGTCATCGCGGAGAGAGAGGTGCTCATGAT 1783  
 Qy 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyArgGlyGlyArgGly 600  
 Db 1784 CGTGTGTCTCCGGTGAAGTGTTCAGAGGTGGCCGTGTGAGACAGAGGTGGCTTCGT 1843  
 Qy 601 G1YGLYArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 620  
 Db 1844 GGTGGCCGGGAGCAGACCGAGTGGCTTGTGTGAGAGAGACAGGTGGCCCTGGGGGG 1903  
 Qy 621 ProProGlyProLeuMetGluGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640  
 Db 1904 CCCCCTGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 Qy 641 LysMetAspLysGlyGlyGlyArgGlyGlyArgGlyGlyArgGlyGlyArgGlyGly 666  
 Db 1964 AAAATGGATAAAGCGAGCAGCCCTCAGAGCGAGAGATCGGCCCTTAC 2011

## RESULT 3

US-09-873-319-717  
 ; Sequence 717, Application US/09873319A  
 ; Publication No. US20030134324A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mungier, William E.  
 ; APPLICANT: Kulkarni, Prakash  
 ; APPLICANT: Geitzenberg, Robert H.  
 ; APPLICANT: Waga, Iwao  
 ; APPLICANT: Yamamoto, Jun  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
 ; FILE REFERENCE: 44921-5029-US  
 ; CURRENT APPLICATION NUMBER: US/09/873,319A  
 ; EARLIER FILING DATE: 2001-06-05  
 ; EARLIER APPLICATION NUMBER: US 60/223,323  
 ; NUMBER OF SEQ ID NOS: 755  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 717  
 ; LENGTH: 2390  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899  
 ; US-09-873-319-717

## Alignment Scores:

Pred. No.: 3.66e-253 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-10-791-017a-2 (1-656) x US-09-873-319-717 (1-2390)

Qy 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 Db 44 ATGGCGCTCAAGATTAAGTACTTAAGCCAGCTGACGCGCAGCGGCTACAGTGTCT 103  
 Qy 21 TyrThrAlaGlnProThrglnGlyTyrAlaGlnThrThrglnAlaTyrGlyGlnGlnSer 40  
 Db 104 TACACCGCCAGCCCACTCAAGATATGACAGACCAACCGGATATGGCAACAAAGC 163  
 Qy 41 TyrGlyThrTyrGlyGlnProThraAspValSerTyrThrglnAlaGlnThrThraAlaThr 60  
 Db 164 TATGGAACCTATGACAGCCCACTGATGTCACTTATACCAAGGCTCAGACCACTGCAACC 223

Qy 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrglnTyrThrnPro 80  
 Db 224 TATGGCAGACCGCCCTATGCAACTTATGAGACAGCTCCACTGTATATATCTCA 283  
 Qy 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrglnAlaTyrAspThr 100  
 Db 284 ACTGCCCCAGGACATACAGCCAGCTGTCCAGGGGTATGCGACTGTATGATACC 343  
 Qy 101 ThrThrAlaThrValThrThrThrglnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 Db 344 ACCACGTCTACAGTCAACACACAGGCTCTTATGACGCTCAGTCTGATATGGACT 403  
 Qy 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThraProGln 140  
 Db 404 CAGCTGCTTATACACCTATGAGGAGAGCCAGCCAGCCAGCCACTACCTACAAACCGCAG 463  
 Qy 141 AspGlyAsnLysProThrglnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160  
 Db 464 GATGGAAACAGCCCACTGAGACTACTCACTCAATCTTACACAGAGGGGTTCACACAG 523  
 Qy 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 Db 524 CCAAGCTTATGATATGACAGATTAACATCACTTATCCAGGTACTTGGAGCTACCCC 583  
 Qy 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrglnPro 200  
 Db 584 ATGCACCCAGTCACTGACCTCATCTTACCTTCCACAGCATTCCTTACACAGCCG 643  
 Qy 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnProSerSerTyr 220  
 Db 644 ACTAGTTATGATCAGACGAGTACTCTCAGCAGAACCACTATGGGACACGACCTAT 703  
 Qy 221 GlnGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 Db 704 GACACAGAGATGCTATGATGATCAACAGAGATATGGAGAGAGCTCCACTTACTTAC 763  
 Qy 241 ProProGlnThrcLysSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 764 CACCCCAAACTGGATCTTACAGCCCAAGCTCCAACTATATGCCAAGACAGCAGCAGC 823  
 Qy 261 TyrGlyGlnGlnSerSerPheArgGlnAspPheProSerSerMetGlyValTyrGlyGln 280  
 Db 824 TACGGGACACAGATTCATTCGACAGAGACCAACCCAGATGATGGGTGTTATGGGCG 883  
 Qy 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 884 GAGCTTGAGAGATTTCCGAGCCAGAGAGAACCGAGCATGAGTGGCCCTGATTAACCG 943  
 Qy 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 Db 944 GCGAGGGGAAGAGGGGATTTGATCGTGAAGCATGACAGAGTGGGGGGGAGAGGA 1003  
 Qy 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyPromet 340  
 Db 1004 CGCGGTGATATGGCGCCCTGAGAGCCAGGAGGCTTCAATTAAGCTGTGAGACCAATG 1063  
 Qy 341 AspGluGlyProAspLysAspLeuGlyProProValAspProAspGlyLysAspAspAsn 360  
 Db 1064 GATGAAGACCAAGATTTATCTTATGAGCCCTCTGTGATCCAGATAGACCTTGAACAC 1123  
 Qy 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCAATTTATGATCAAGATTAATTAATGACAGTGTACTGTATGATCTGGCAGACTTC 1183  
 Qy 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 Db 1184 TTAAACAGTGTGGGCTTGTAAATGAACAGAGAACTGGGCAACCAATGATCCACATC 1243  
 Qy 401 TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysAspPro 420  
 Db 1244 TACCTGACAGAGAAACAGAAAGCCCAAGGAGGAGCCACAGTGTCTTATGAGAACCCA 1303  
 Qy 421 ProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerLysLeu 440

```

Db      1304 CCCACTGCCAGCTGCGCTGGAATGCTTGAATGGAAGATTTTCAAGGAGCAAACTT 1363
Qy      441 LysValSerLeuAlaArgLysLysProPheMetAsnSerMetArgLysLysLeuProPro 460
Db      1364 AAAGTCTCCCTGCTCGGAAGAAGCTCCATGAAAGATATGGGGGTGGTCTGCCAACC 1423
Qy      461 ArgGluLysArgLysMetProProProLeuArgLysLysProLysLysProProPro 480
Db      1424 CGTAGGAGGAGAGAGCAAGCAACCACTCGTAGAGGTCCAGAGAGCCAGAGAGGTCT 1483
Qy      481 G1YGLYProMetG1YArgMetG1YArgG1YGLYAspArgG1YGLYPheProProArg 500
Db      1484 GGGGAGACCATGGGTCCCATGGAGGCGCTGAGAGAGATAGAGAGGCTTCTCCCAAGA 1543
Qy      501 G1YProArgLysSerArgLysAsnProSerG1YGLYAsnValGlnH1AsrArgLag1Y 520
Db      1544 GGAACCCGGGGTTCGCCAGAGGAAACCTCTGAGAGAGAAACGTCCAGCAACGAGCTGGA 1603
Qy      521 AspTyrGlnCybProAsnProG1YCybG1YAsnGlnAsnPheAlaTyrArgThrGluCyb 540
Db      1604 GACTGGCAGTGTCCCATCCGGGTGTGGAAACCAAGAACTTCGCTGAGAAACAGAGTGC 1663
Qy      541 AsnGlnCybLysAlaProLysProGlnG1YPheLeuProProProPheProProG1Y 560
Db      1664 AACCAAGTAAAGGCCCAAGAGCTTGAAGGCTTCTCCGCAACCTTCCGGCCCGGGGT 1723
Qy      561 G1YAspArgLysArgLysG1YProG1YGLYMetArgG1YGLYArgG1YGLYLeuMetAsp 580
Db      1724 GGTGATCGTGGCAGAGGTGGCCCTGTGGCATGGGGAGAGAGAGGTGGCTCATGGAT 1783
Qy      581 ArgG1YGLYProG1YGLYMetPheArgG1YGLYArgG1YGLYAspArgG1YGLYPheArg 600
Db      1784 CGGTGGTGTCCCGTGAAGTTCAGAGGTGGCGGTGTGAGACAGAGGTGGCTTCCT 1843
Qy      601 G1YGLYArgG1YMetAspArgG1YGLYPheG1YGLYGLYArgArgG1YGLYProG1YGLY 620
Db      1844 GGTGGCGGGGCGCATGACCAAGGTGCTTGTGGAGAGAGAGAGAGGTGGCTTGGGGGG 1903
Qy      621 ProProG1YProLeuMetGlnMetG1YGLYArgArgG1YGLYArgG1YGLYProG1Y 640
Db      1904 CCCCCTGGAACCTTGTATGAAACAGATGGAGAGAGAGAGAGAGGTGGAGGACCTTGA 1963
Qy      641 LysMetAspLysG1YGLYAsnArgGlnGlnArgArgAspArgProTyr 656
Db      1964 AAATGATAAAGCGAGCAACCTGCAAGAGAGAGAGAGAGGTGGCCCTTAC 2011

```

RESULT 4

US-10-791-017A-1  
 / Sequence 1, Application US/10791017A  
 / Publication No. US20040197827A1  
 / GENERAL INFORMATION:  
 / APPLICANT: JENAPHARM GmbH & Co. KG  
 / TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances  
 / FILE REFERENCE: Pat 3684/11  
 / CURRENT APPLICATION NUMBER: US/10/791.017A  
 / NUMBER OF SEQ ID NOS: 7  
 / SOFTWARE: PatentIn Ver. 2.1  
 / SEQ ID NO 1  
 / LENGTH: 2390  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: (44)..(2011)  
 / OTHER INFORMATION: EMS  
 / US-10-791-017A-1

Alignment Scores:  
 Pred. No.: 3.66e-253 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0
US-10-791-017A-2 (1-656) x US-10-791-017A-1 (1-2390)
Qy      1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrSerAla 20
Db      44 ATGGCGTCCACGGATTAACATGACTTAAGCCAAAGCTGACGGAGCGAGGCTCAAGTGT 103
Qy      21 TyrThrAlaGlnProThrGlnGlnTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 40
Db      104 TACACCGCCCAAGCCCACTCAAGATATGACACACACCCAGGCAATATGGGCAACAAAGC 163
Qy      41 TyrG1YThrTyrG1YglnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60
Db      164 TATGAACCTATGAGACAGCCCACTGATGTGAGTATACCAAGGCTCAGACCACTGCAAC 223
Qy      61 TyrG1YglnThrAlaTyrAlaThrSerTyrG1YglnProProThrG1YTYrThrThrPro 80
Db      224 TATGGGCAAGCCGCTATGCAACTTCTTATGACAGGCTCCCACTGGTATATCACTCA 283
Qy      81 ThrAlaProGlnAlaTyrSerGlnProValGlnG1YTYrG1YThrG1YAlaTyrAspThr 100
Db      284 ACTGCCCCCAGGCAATACAGCCAGGCTGTCCAGGGTATAGCACTGGTCTTATGATACC 343
Qy      101 ThrThrAlaThrValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrG1YThr 120
Db      344 ACCACTGTACATGACACACCAAGGCTCTTATGAGCTCAGTCTGATATGGCACT 403
Qy      121 GlnProAlaTyrProAlaTyrG1YglnGlnProAlaAlaThrAlaProThrArgProGln 140
Db      404 CAGCTGCTTATCAGCTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
Qy      141 AspG1YAsnLysProThrGlnThrSerGlnProGlnSerSerThrG1YTYrAsnGln 160
Db      464 GATGAAACAGGCCCACTGAGACTTATGCAACTCATATGACACAGGGGTATCAACAG 523
Qy      161 ProSerLeuG1YTYrG1YglnSerAsnTyrSerTyrProGlnValProG1YSerTyrPro 180
Db      524 CCCAGCTAGATATGAGACAGAGTAACTACAGTATACCCAGGTAACCTGGAGAGCTACCCC 583
Qy      181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db      584 ATGACAGCAAGTCACTCACTTCATCCCTCAACAGCTATTCCTTACACAGCCG 643
Qy      201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrG1YglnProSerSerTyr 220
Db      644 ACTAGTTATGATCAGAGCACTTCTTCAGCAAGAACCTTATGGGCAACCGAGCACTAT 703
Qy      221 G1YglnGlnSerSerTyrG1YglnGlnSerSerTyrG1YglnGlnProProThrSerTyr 240
Db      704 GAGACAGAGTATGATGATGATCAACAAAGCAGTATAGGAGAGAGAGAGAGAGAGAGAG 763
Qy      241 ProProGlnThrG1YSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260
Db      764 CCAACCCCAACCTGATCTTACAGCAAGCTCCAAAGTCAATATGCAACAGAGAGAGAG 823
Qy      261 TyrG1YglnGlnSerSerPheArgGlnAspH1AspSerSerSerMetG1YValTyrG1Ygln 280
Db      824 TACGGGAGAGAGTTCATTTCCAGAGAGACCAACCAAGTATGAGTGTATTTATGGGAG 883
Qy      281 GlnSerG1YGLYPheSerG1YProG1YAsnArgSerMetSerG1YProAspAsnArg 300
Db      884 GATGCTGAGAGATTTTCCGAGCAGAGAGAGAAACCGAGCATGAGTGCCCTGATTAACCG 943
Qy      301 G1YArgG1YArgG1YGLYPheAspArgG1YGLYMetSerArgG1YGLYArgG1YGLY 320
Db      944 GGCAGGGAGAGAGAGAGATTTGATCGTAGAGCATAGAGAGAGTGGCGGGAGAGAGAGA 1003
Qy      321 ArgG1YGLYMetG1YSerAlaG1YArgG1YGLYPheAsnLysProG1YGLYProMet 340
Db      1004 CGCGGTGGAATGGGACGCGTGAAGAGAGAGAGAGAGTGGCTTCAATAGCTGTGTGAACCAT 1063

```

```

QY 341 AspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn 360
Db 1064 GATGAAAGACCAATCTTGAATCTAGGCGCCACTTGAATCAAGTAGAGCTCTGACAC 1123
QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380
Db 1124 AGTGCAATTTATGTACAAAGATTAAATGACAGTGTGACTGATGATGATCTGGGCACTTC 1183
QY 381 PheIysGlnCysGlyValValIleValIleMetAsnIleYargThrGlyGlnProMetIleHisIle 400
Db 1184 TTTAAGAGGTGGGTGTGTATGATGAAACAAAGAACTGGGCAACCAATGATCCACATC 1243
QY 401 TyrLeuAspIysGluThrGlyIleProIysGlyAspAlaThrValIleSerTyrGluAspPro 420
Db 1244 TACCTGACAAAGAAACAGAAAGCCAAAGCCATCCACAGTGTCTTGAAGACCA 1303
QY 421 ProThrAlaIysAlaAlaValGluTyrPheAspGlyIleYargAspPheGlnGlySerIleVal 440
Db 1304 CCCACTGCCAAGGCTCCGTGGAATGTTGATGGAAGATTTTCAAGGAGCAAACTT 1363
QY 441 LysValSerLeuAlaGlyIleIysProProMetAsnSerMetArgGlyIleLeuProPro 460
Db 1364 AAAGTCTCCCTTGTCTCGAAGAAACCTCCAAATGAAAGATGCGGGGTGTGTGCCACCC 1423
QY 461 ArgGluGlyIysArgIysMetProProProLeuAspGlyIleProIysGlyIlePro 480
Db 1424 CGTAGAGGAGAGGACATGCAACCACTCCGTGAGAGTCCAGAGGCCCAAGAGTCTCT 1483
QY 481 GlyIleProMetGlyIysArgMetGlyIleYargGlyIleYargAspGlyIlePheProProArg 500
Db 1484 GGGGGAACCAATGGGTCCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCCAAAG 1543
QY 501 GlyProArgGlySerArgGlyIysAsnProSerGlyIleYargValGlnIleAspAlaGly 520
Db 1544 GGAACCCGGGGTCTCCAGAGAAACCTCTGAGAGAGAAACGTCAGACCGAGCTGGA 1603
QY 521 AspTyrGlnCysProAsnProGlyIleYargIleYargIleYargIleYargIleYarg 540
Db 1604 GACTGGCAGTGTCCCATTCGGGTGTGGAACCAAGACTTCCGCTGGAAGACAGAGTCC 1663
QY 541 AsnGlnCysIleYargIleYargIleYargIleYargIleYargIleYargIleYarg 560
Db 1664 AACCAAGTAAAGGCCCAAGAGCTGAAAGCTTCCCTCCGCGCACCTTCCGCGCGGT 1723
QY 561 GlyAspArgGlyIysArgGlyIleYargIleYargIleYargIleYargIleYarg 580
Db 1724 GGTGATGTGGCAGAGGTGGCTGTGTGGCATGCGGGAGAGAAAGGTGGCTTCATGAT 1783
QY 581 ArgGlyIleProGlyIleYargIleYargIleYargIleYargIleYargIleYarg 600
Db 1784 CGTGTGTGTCCCGGTGGAATTTCAAGAGTGGCCGTGTGGAAGCAAGAGTGTTCCT 1843
QY 601 GlyIleYargIleYargIleYargIleYargIleYargIleYargIleYargIleYarg 620
Db 1844 GGTGGCCGGGAGATGACCGAGGTGTGTGTGTGAGAAAGAGAGAGAGAGAGAGAGAG 1903
QY 621 ProProGlyProLeuMetGlnIleMetGlyIleYargIleYargIleYargIleYarg 640
Db 1904 CCCCTGACCTTTGATGAGAAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1963
QY 641 LysMetAspIysGlyIleYargIleYargIleYargIleYargIleYargIleYarg 656
Db 1964 AAAATGATAAAGCCAGACCCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011

```

```

APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakar
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822, 830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195, 604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 2176
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-49

Alignment Scores:
Pred. No.: 3,53e-251 Length: 2176
Score: 3605.00 Matches: 655
Percent Similarity: 99.85% Conservative: 0
Best Local Similarity: 99.85% Mismatches: 1
Query Match: 99.23% Indels: 1
DB: 9 Gaps: 0

US-10-791-017A-2 (1-656) x US-09-822-830A-49 (1-2176)
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20
Db 2160 ATGGCGTCCACCGGATTACAGTACCTATACCAAGTGCAGCGGAGGAGGCTACAGTCT 2101
QY 21 TyrThrAlaGlnProThrGlnGlnGlyTyrAlaGlnThrThrGlnAlaIleYargIleGlnSer 40
Db 2100 TACACCGCCAGCGCCACTCAAGATATGCAAGACCAACCAAGCATATGGCAACAAAGC 2041
QY 41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
Db 2040 TATGAACTATGAGACAGCCCACTATATGTCAGTATACCAAGCTCAGACCTGCAACC 1981
QY 61 TyrGlyGlnThrAlaIleYargIleYargIleYargIleYargIleYargIleYarg 80
Db 1980 TATGGCAGACCGCCTATGCACTTCTTATGAGACAGCTCCCACTGTTATATCTCA 1921
QY 81 ThrAlaProGlnAlaIleYargIleYargIleYargIleYargIleYargIleYarg 100
Db 1920 ACTGCCCCAGGACATACAGCCAGCTGTCCAGGGGTATGGCACTGTGTCTTATGATACC 1861
QY 101 ThrThrAlaThrValIleThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
Db 1860 ACCACTGTACACTCACACCAACCGGCTCTTATGAGCTCAGTGTGATATGGCACT 1801
QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140
Db 1800 CAGCGCTTATACAGCTATGAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1742
QY 141 AspGlyAsnIleYargIleYargIleYargIleYargIleYargIleYargIleYarg 160
Db 1741 GATGAAACAAAGCCCACTAGACTCACTCACTCACTCACTCACTCACTCACTCACTCA 1682
QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProIysSerTyrPro 180
Db 1681 CCCAGCTTAGATATGAGAGAGTAACTCAAGTATATCCCAAGTATCTGAGACTACCCC 1622
QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db 1621 ATGCAGACGATCACTGCACTCATCTTACCTCCACAGCATATCTCTTACAGAGCGG 1562
QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnSerThrTyrGlyGlnProSerSerTyr 220
Db 1561 ACTAGTTATGATCAGAGCAGTACTCTCAGAGAAACCACTATGGGAGCAACCGAGAGCTAT 1502

```

```

RESULT 5
US-09-822-830A-49/c
; Sequence 49, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary

```

```

QY 221 G1yGlnGlnSerSerTyrg1yGlnGlnSerSerTyrg1yGlnGlnProProThrSerTyrg1y 240
Db 1501 GGACAGACAGAGTGTGCTTGTGCAACAAAGAGCTATGAGCGAGCGCTCCACAGTGTAC 1442
QY 241 ProProGlnThrg1ySerTyrg1ySerGlnAlaProSerGlnThrg1ySerGlnGlnSerSer 260
Db 1441 CCACCCCAAACTGGATCTTACAGCCAAAGCTCCAACTCAATATACCAACAGAGAGAGC 1382
QY 261 Tyrg1yGlnGlnSerSerPheArgGlnAph1aProSerSerMetG1yAlaTyrg1yGln 280
Db 1381 TACGGGAGAGAGAGTATTCCTCCAGAGACCACTCCAGTACATGGGTGTATGGGAG 1322
QY 281 GluSerG1yG1yPheSerG1yProG1yG1yLysAph1aArgSerSerG1yProAphAphArg 300
Db 1321 GAGCTGTGAGAGATTTTCCGAGCCAGAGAGAAACCGAGCATGATGCGCTCGATTAACCGG 1262
QY 301 G1yArgG1yArgG1yG1yPheAphArgG1yG1yMetSerArgG1yG1yArgG1yG1y 320
Db 1261 GGCAAGGAGAGAGAGAGAGATTTGATCTGTGAGGAGCATGAGAGAGGTGGCGGAGAGAG 1202
QY 321 ArgG1yG1yMetG1ySerArgG1yG1yLysArgG1yG1yPheAphArgProG1yG1yProMet 340
Db 1201 CGCGGTGAAATGGGAGAGCGGTGAGAGAGAGGTGCTTCAATAGCCGTGGTGAACCATG 1142
QY 341 ArgG1yG1yProAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 360
Db 1141 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082
QY 361 SerAla1a1eTyrg1yGlnGlnLysAphAphAphAphAphAphAphAphAphAphAphAph 380
Db 1081 AGTGCAATTAATGATGACAGAGATTAATGACAGTGTGATCTGATGATGATCTGAGACTTC 1022
QY 381 PheLysGlnCysG1yVal1yVal1yMetAphAphAphAphAphAphAphAphAphAphAph 400
Db 1021 TTTAAGAGAGTGGGAGTGTGTTAATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 962
QY 401 Tyrg1yLysAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 420
Db 961 TACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
QY 421 ProThra1a1yVal1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1 440
Db 901 CCCACTGCAAGAGTGTGCGGTGAGAGAGTGTGATGAGAGAGATTTTCAAGGAGAGAACTT 842
QY 441 LysValSerLeuAlaArg1yLysProProMetAphAphAphAphAphAphAphAphAphAph 460
Db 841 AAAGTCTCCCTGTGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
QY 461 ArgG1yGlnG1yMetProProProProLeuAphAphAphAphAphAphAphAphAphAph 480
Db 781 CGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
QY 481 G1yG1yProMetG1yArgMetG1yG1yArgG1yG1yLysAphAphAphAphAphAphAphAph 500
Db 721 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY 501 G1yProArgG1ySerArgG1yAphAphAphAphAphAphAphAphAphAphAphAphAph 520
Db 661 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
QY 521 AapTPrGlnCysProAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 540
Db 601 GACTGGAGAGTGTCCCAATCCGGGTGTGAGAAACCAAGAACTTCCGTGAGAGAGAGAG 542
QY 541 AaaglnCysValAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 560
Db 541 AACCAGAGTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
QY 561 G1yAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 580
Db 481 GGTATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422

```

```

QY 581 ArgG1yG1yProG1yG1yMetPheArgG1yG1yArgG1yG1yAphAphAphAphAphAph 600
Db 421 CGTGTGTGTCCCGATGAGAACTTCAAGAGTGTGCGGTGTGAGAGAGAGAGAGAGAGAG 362
QY 601 G1yG1yArgG1yMetAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 620
Db 361 GTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
QY 621 ProProG1yProLeuMetGlnGlnMetG1yG1yArgAphAphAphAphAphAphAphAph 640
Db 301 CCCCTGTGAGCTTGTGATGAGAAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
QY 641 LysMetAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 656
Db 241 AAAATGATTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194

RESULT 6
US-09-822-830A-410/c
: Sequence 410. Application US/09822830A
: Patent No. US20020142952A1
: GENERAL INFORMATION:
: APPLICANT: Genetics Institute, Inc.
: APPLICANT: Mong, Gordon G.
: APPLICANT: Clark, Hilary
: APPLICANT: Fectel, Kim
: APPLICANT: Agostino, Michael J.
: APPLICANT: Howes, Steven H.
: APPLICANT: Resnick, Richard J.
: APPLICANT: Gulukota, Kamalakart
: APPLICANT: Graham, James R.
: TITLE OR INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
: FILE REFERENCE: GIN 6402
: CURRENT APPLICATION NUMBER: US/09/822,830A
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/195,604
: PRIOR FILING DATE: 2000-04-06
: NUMBER OF SEQ ID NOS: 631
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 410
: LENGTH: 2273
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-822-830A-410

Alignment Scores:
Pred. No.: 2,36e-242 Length: 2273
Score: 3483.00 Matches: 636
Percent Similarity: 97.10% Conservative: 0
Best Local Similarity: 97.10% Mismatches: 1
Query Match: 95.87% Indels: 18
DB: Gaps: 2

US-10-791-017A-2 (1-656) x US-09-822-830A-410 (1-2273)
QY 2 AlasSerThrAapTyrg1ySerThrTyrg1ySerGlnAlaAlaGlnGlnGlnTyrg1ySerAlaTyrg1 21
Db 2263 GCGTCCACGAGATTACAGTATACCTTAAGCAAGCTGCAAGCGGAGAGAGAGAGAGAGAG 2204
QY 22 ThrAlaGlnProThrg1yGlnGlnTyrg1yAlaGlnThrThrg1yAlaTyrg1yGlnGlnSerTyrg1 41
Db 2203 ACCGCCAGGCCCACTCAAGATATGACAGACCAACCAAGCATATGAGGCAAAAGCTAT 2144
QY 42 G1yThrTyrg1yGlnProThraAphAphAphAphAphAphAphAphAphAphAphAphAphAph 61
Db 2143 GGAACCTATGAGACGCCCACTGATGTCACTATACCAAGGCTCAGACCACTGCAACTAT 2084
QY 62 G1yGlnThra1a1yThra1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1 81
Db 2083 GGGCAGAGCGCCATATGAGAACTTCTTAAGACAGCTCCCACTGGTTATTACTACTCAACT 2024
QY 82 AlaProGlnAlaTyrg1ySerGlnProValGlnG1yTyrg1yThrg1yAlaTyrg1yThra1a1a1 101
Db 2023 GCCCCCAAGGAGATACAGCCAGCTGTCCAGGGGATATGCACTGTGTCTTATGATACCAAC 1964

```

QY 102 ThrAlaThrValThrThrThrGlnAlaSerTyrAlaIleGlnSerAlaTyrGlyThrGln 121  
 Db 1963 AGCTCAAGACACACACACACAGCCCTCTATGAGTCACTGATGATGAGCACTCAG 1904  
 QY 122 ProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGlnAsp 141  
 Db 1903 CTTGCTTATCCAGCTTATGGGCAACAGACAGACCACTGACCTTACAAAGCCGAGAT 1844  
 QY 142 GlyAlaAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGlnPro 161  
 Db 1843 GGAACCAAGCCACTGAGACTGATCAACCTCAATCTGACACAGGGGCTTCAACCAACCCC 1784  
 QY 162 SerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrProMet 181  
 Db 1783 AGCTTAAAGATATGACAGAGTACTCACTGATTTATCCAGGATCTGGAGCTTACCCATG 1724  
 QY 182 GlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThr 201  
 Db 1723 CAGCCAGTCACTGACCTCCATCTTACCTTCAAGCTATTTCTTACACAGCCGACT 1664  
 QY 202 SerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyrGly 221  
 Db 1663 AGTATGATCAGACAGATTACTCTCAGACAGAACCTATGGGCAACGAGCAGCTATGGA 1604  
 QY 222 GlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyrPro 241  
 Db 1603 CAGCAGATGACTTATGCTCAACAAAGACCTTATGGGAGAGCCCTCCCATGTTATCCCA 1544  
 QY 242 ProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerTyr 261  
 Db 1543 CCCCAGTGGATCTTACAGCCAGCTCCAAATGATGACCAACAGAGAGCAGTAC 1484  
 QY 262 GlyGlnGlnSerSerPheAsnGlnAspHisProSerSerMetGlyValTyrGlyGlnGlu 281  
 Db 1483 GGGGAGAGAGTTATTCGACAGAGACACCCAGTGCATGGGTGTTTATGGGCAAGAG 1424  
 QY 282 SerGlyGlyPheSerGlyProGlyGlyAsnArgSerSerSerGlyProAspAsnArgGly 301  
 Db 1423 TCTGAGAGATTTTCCGAGCCAGAGAGAACCCAGCATGAGTGGCCCTGATTAACCGGAGGC 1364  
 QY 302 ArgGlyArgGlyGlyPheAsnArgGlyGlyMetSerArgGlyGlyArgGlyGlyArg 321  
 Db 1363 AGGGGAGAGAGGGGGATTTGATCTGAGGACATGAGAGAGTGGCGGGAGAGAGCGC 1304  
 QY 322 GlyGlyMetGlySerAlaGlyGlyArgGlyGlyPheAsnLysProGlyGlyProMetAsp 341  
 Db 1303 GGTGGAATGGGC--GCTGAGAGACGAGGTGGCTTCAATAAGCTGTGAGCCCATGAT 1247  
 QY 342 GlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsnSer 361  
 Db 1246 GAGAGACCAATCTTGAATCTTAGGCCCACTGTAGATCCAGATGAAAGCTGTGACACAGT 1187  
 QY 362 AlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 381  
 Db 1186 GCAATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGACTTCTT 1127  
 QY 382 LysGlnLysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIleTyr 401  
 Db 1126 AAGCAGTGTGGGTGTTGATGATGACAAAGAACTGGGCAACCAATGATCCACTTAC 1067  
 QY 402 LeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlyAspProPro 421  
 Db 1066 CTGGACAAAGAAACAGAAAGCCCAAGGCGATTCACAGTGTCTTATGAAAGCCCAACC 1007  
 QY 422 ThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeuLys 441  
 Db 1006 ACTGCCAAGGCTGCGTGAATGTTGTTGGGAAAGATTTCAGAGGAGCAAACTTAA 947  
 QY 442 ValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProProArg 461  
 Db 946 GTCTCCCTTCTCGAAGAAAGCTCCCAATGAACGATGCGGGGTGTCTGCAACCCCGT 887

QY 462 GlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyProGly 481  
 Db 886 GAGGGCAAGGATGCAACCACTCCGTTGAGGTCCAGAGAGCCCAAGAGGTCTTGGG 827  
 QY 482 GlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArgGly 501  
 Db 826 GAGCCATGAGGTGCGATGGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGA 767  
 QY 502 ProArgLysSerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisAlaGlyAsp 521  
 Db 766 CCCCAGGCTTCCGAGAGAAACCCCTCTGAGAGAGAAACCTCCAGACCCAGGTGGAGAC 707  
 QY 522 TyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaThrPheThrGlnCysAsn 541  
 Db 706 TGGCAGTGTCCCAATCCGGTTGTGAAACCAAACTTCCCTGAGAAACAGAGTGCAC 647  
 QY 542 GlnCysLeuAlaProLysProGlnGlyPheLeuProProProPheProProGlyGly 561  
 Db 646 CAGTGT-----GGT 638  
 QY 562 AspArgGlyArgGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAspArg 581  
 Db 637 GATCGTGGCAGAGTGGCCCTGTGTGCAATGCGGGAGAGAGAGTGGCCCATGAGATCGT 578  
 QY 582 GlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArgGly 601  
 Db 577 GGTGTCCCGGTGGATGTTTCAAGGTGGCCGTGTGTGAGACAGAGTGGCTTCCGTGT 518  
 QY 602 GlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGlyPro 621  
 Db 517 GCGCGGGGATGAGACCGAGGTGGCTTGTGTGAGAGAAAGAGAGTGGCCCTGGGGGGCCC 458  
 QY 622 ProGlyProLeuMetGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGlyGly 641  
 Db 457 CCGTGAACCTTGTATGGAACAGATGGAGAGAAAGAGAGAGTGGAGGACCTTGAAAA 398  
 QY 642 MetAspLysGlyGlyLysArgGlnGluArgArgAspArgProTyr 656  
 Db 397 ATGATTAAGGCGAGCACCTCTCAGAGCCGAGAGATCGGCCCTTAC 353

RESULT 7  
 US-10-094-749-1074  
 ; Sequence 1074, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHICO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094, 749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350, 435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 1074
; LENGTH: 1988
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1074

Alignment Scores:
Pred. No.: 1,55e-227      Length: 1988
Score: 3277.00           Matches: 599
Percent Similarity: 91.31%  Conservative: 0
Best Local Similarity: 91.31%  Mismatches: 1
Query Match: 90.20%       Indels: 56
DB: 17                    Gaps: 1

US-10-791-017a-2 (1-656) x US-10-094-749-1074 (1-1988)

Qy      1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20
Db      12 ATGGCGTCCACGGATTACAGTACTTATGCGCAAGCTGACGCGACGGGCTACAGTGT 71
Qy      21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlyGlnGlnSer 40
Db      72 TACACGCGCCACGCGCACTCAAGGATATGACACAGCACCGGCAATGGGCAACAAAGC 131
Qy      41 TyrGlyTyrThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
Db      132 TAGGAACCTATGACAGCCCACTGATGTCAGCTATACCCAGGCTCAGACCTCAGAAC 191
Qy      61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80
Db      192 TAGGGACAGACCCCTATGCACTTCTTATGACAGCTCCCACTGGTTATCTACTCCA 251
Qy      81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyTyrGlyAlaTyrAspThr 100
Db      252 ACTGCCCGCCAGCATACAGCCAGCTGTCCAGGGGATGACACTGGTGTATGATACC 311
Qy      101 ThrThrAlaThrAlaThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
Db      312 ACCACGCTACAGATCACACACACACAGGCTCTTATGCACTGATGCTGATATGGCACT 371
Qy      121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140
Db      372 CAGCGCTTATTCAGGCTTATGGGACAGCCAGCCAGCCACTGCA----- 416
Qy      141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160
Db      416 ----- 416
Qy      161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
Db      416 ----- 416
Qy      181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db      417 -----CTTACAGACTATTCCTTACACAGCCG 443
Qy      201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220
Db      444 ACTAGTTATGATCAGAGCAGTACTCTCAGCAGAACCTATGGGCAACCGACAGACTAT 503
Qy      221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db      504 GGCAGACAGGGTATGCTATGCTCAACAAAGCAGCTATGGGCAAGGCTCCCACTAGTTAC 563
Qy      241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260
Db      564 CCACCCCAAACTGATCTTACAGCCAAAGCTCCAAAGTCAATATGACCAACAGACAGACAC 623
Qy      261 TyrGlyGlnGlnSerSerPheArgGlnAspAlaProSerSerMetGlyAlaTyrGlyGln 280
Db      624 TAGGGCAGAGAGATTCAATCCGACAGGACCAACCCCAAGTACATGGGTGTTTATGGGCA 683
Qy      281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300

```

```

Db      684 GAGCTCGAGAGATTTCGACCAAGAGAAACCGGACATGATGACCTTGATTAACCGG 743
Qy      301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyAlaArgGlyGly 320
Db      744 GGCAGGGGAAGAGGGGGATTGATTCGTGAGGATATGACAGAGGTGGCGGGGAGAGGA 803
Qy      321 ArgGlyGlyMetGlySerAlaGlyGlyAlaArgGlyGlyPheAsnLysProGlyGlyProMet 340
Db      804 CGCGGTGATGATGGGACGCGCTGAGAGCGAGGCTGCTTCAATTAAGCTCGTGGACCATG 863
Qy      341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360
Db      864 GATGAAGACCAATCTTATGATTAAGGCCCACTGTGATCAATGAAGAAGCTTGACAC 923
Qy      361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrAspAspLeuAlaAspPhe 380
Db      924 AGTGCATTTTATTTATCAAGATTTAATGACAGTGACTTATGATGATCTGGCAGACTTC 983
Qy      381 PheGlyGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
Db      984 TTTAAGCAGTGTGGGTTGTTAAGATGAAACAAAGAACTGGCAACCATGATCCACATC 1043
Qy      401 TyrLeuAspLysGlyLuthrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420
Db      1044 TACCTGACAAAGAAACAGGAAAGCCAAAGCGATGCCACAGTGTCTATGAAAGACCA 1103
Qy      421 ProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerLysLeu 440
Db      1104 CCACAGTCCAAAGGCTCCGTGGAATGTTGATGGGAAGATTTTCAAGGAGCAAACTT 1163
Qy      441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProPro 460
Db      1164 AAGTCTCCCTTCTCGGAAAGACCTCCATGAACAGTATGGCGGTGTCTGCGCACCC 1223
Qy      461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro 480
Db      1224 GTGAGGGACAGAGGCAATGCCACCACTCCGTGAGGTCCAGAGGCCCAAGAGGTCT 1283
Qy      481 GlyLysProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500
Db      1284 GGGGAGACCATGGGTGTGCATGGAGGCCGTGAGAGATGAGAGAGGCTTCCCTCCAGA 1343
Qy      501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520
Db      1344 GGAACCCGGGGTCCGAGAGGAACCTCTGAGAGAGAAACGTCCAGACCAAGAGCTGGA 1403
Qy      521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThrGlyCys 540
Db      1404 GACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCGCTGGAGAAACAGAGTGC 1463
Qy      541 AsnGlnCysLeuAlaProLysProGlyGlyPheLeuProProProPheProProGly 560
Db      1464 AACCACTGTAAGGCCCAAAAGCTGAAGGCTTCTCCGCACTTTCGCCCCGGGT 1523
Qy      561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580
Db      1524 GGTGATCTGGGACAGAGTGGCCCTGGTGGCATCGGGGAGAGAGGTGGCTCATATGAT 1583
Qy      581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600
Db      1584 CGTGTGTCTCCGTGTGAATGTTCAAGGTGGCGGTGTGGAGAACAGAGGTGCTTCCGT 1643
Qy      601 GlyLysArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620
Db      1644 GGTGGCGGGGCAATGACCAAGTGTCTTGTGTGAGAGAAAGAGAGTGGCTCGGGGGG 1703
Qy      621 ProProGlyProLeuMetGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640
Db      1704 CCCCTGACACTTTGATGAAACAGATGGGAGAAAGAGAGAGAGCTGAGAGACCTGGA 1763
Qy      641 LysMetAspLysGlyLuthisArgGlnLuthArgAspArgProTyr 656

```

Db 1764 AATAATGATMAAGGCGAGCACCGTCAGAGCGAGAGATCGCCCTAC 1811

RESULT 8

US-10-439-703-58

Sequence 58, Application US/10439703

Publication No. US20040018527A1

GENERAL INFORMATION:

APPLICANT: Chang, Jenny

APPLICANT: O'Connell, Peter

TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax

TITLE OF INVENTION: Chemoresistivity and Chemoresistance

FILE REFERENCE: HO-P24820S1/10205813

CURRENT APPLICATION NUMBER: US/10/439, 703

CURRENT FILING DATE: 2003-05-16

PRIOR APPLICATION NUMBER: US 60/381,141

PRIOR FILING DATE: 2002-05-17

NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn version 3.1

SEQ ID NO 58

LENGTH: 1824

TYPE: DNA

ORGANISM: Human

US-10-439-703-58

Alignment Scores:

Pred. No.:	1,52e-83	Length:	1824
Score:	1282.00	Matches:	301
Percent Similarity:	53.81%	Conservative:	59
Best Local Similarity:	44.99%	Mismatches:	153
Query Match:	35.29%	Indels:	156
		Gaps:	30

US-10-791-017a-2 (1-656) x US-10-439-703-58 (1-1824)

Qy 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20

Db 79 ATGGCCCTCAACAGCATTTATACC-----CAACACCAACCAACCAACCTATGGGGCC 126

Qy 21 TyrThrAlaGlnProThrGlnGlyTyrAla---GlnThrThrGlnAlaTyrGlyGlnGln 39

Db 127 TACCCCAACCCAGCCCGGGGAGGCTATTCAGAGAGAGCATGCGCTTACGAGACAGCAG 186

Qy 40 SerTyrGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAla 59

Db 187 AGTTACAGGTGTTATGCGACGTCACGAC-----ACTTCA 222

Qy 60 ThrTyrGlnThrAlaTyrAlaThrSerTyrGlyGlnProPro---ThrGlyTyrThr 78

Db 223 GGCTATGGCCAGAGAGCTAT--TCTTCTTATGGCCAGAGCAAGACAGGCTATGGA 279

Qy 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrGlyAla 97

Db 280 ACTCAGTCACTCC-----CAGGATATGCTGCTGACTGGGGGC 318

Qy 98 TyrAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAla 117

Db 319 TATGGC-----AGTAGCCAGAGCTCCCATGCTCT 348

Qy 118 TyrGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137

Db 349 TAGGGCAGAGAGCTCTTACCTGCTATGGCCAGAGACAGCT----- 393

Qy 138 ArgProGlnAspGlyAsnLysProThrGluThrSerGlnProGlnSerSerThrGly 157

Db 394 -----CCAGCAGACACTGGGAAGT 414

Qy 158 TyrAsnGlnProSerLeuGlyTyrGlyGlnSerAsnLysSerTyrProGlnValProGly 177

Db 415 TAGCGTAGAGCTCT-----CAGAGCAGAGCATATGGAGCCCGC---AGTGG 462

Qy 178 SerTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSer 197

Db 463 AGCTACAGCCAGAGCT-----AGCTATGTGGA 492

Qy 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnThrThrTyrGlyGlnPro 217

Db 493 -----CAGCAGCAAGCTATGACAGAGCAAGCATTAATCCCT 534

Qy 218 SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProPro 237

Db 535 CAGGCTATGAGACAGCAACAGTACACAGAGCATGCTGTGAGAGTGAAGTGA 594

Qy 238 ThrSerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGln 257

Db 595 GTGTGA-----GGTACTATGCGCAAGATCAATCTTCATGATGATGTGT 639

Qy 258 SerSerSerTyrGlyGlnGlnSerSerPheAspGlnAspHisProSerSerMetGlyVal 277

Db 640 GTGGGACAGGTGCGCGTTATGGCAATCAAGACAGAGTGTGAGGTGACCGGTGGC 699

Qy 278 TyrGlyGlnGlnSerGlyGlyPheSerGlyProGlyGlyLysAspSerMetSerGlyPro 297

Db 700 TATGACAGAGAGACCGTGA-----GCCCGCGCAGGGGTGACATGCT-- 744

Qy 298 AspAsnArgGlyLysGlyLysGlyLysPheAspArg-----GlyGlyMetSerArgGly 315

Db 745 GCGCGCGCGCGCGCGCGCGGTGTTCAACCGCAGCATGTGCTATGAACCCAGA 804

Qy 316 GlyArgGlyGlyLysGlyLysGlyLysMetGlySerAlaGly-----GluArgGlyGlyPhe 333

Db 805 GGTCTGTGAGGTGCGCGTGTGAGGACAGAGGTGCGATGCGGCAAGTACCTGTGTGCTTC 864

Qy 334 AsnLysProGlyLysGlyProMetAspGlnGlyLysProAspLeuAspLeuGlyProProValAsp 353

Db 865 AATTAATTTGTGTGCGCTCGGACCAAGATACCGTATGAC-----TCC 909

Qy 354 ProAspGlnAspSerAspAsnSerAlaLysTyrValGlnGlyLeuAsnAspSerValThr 373

Db 910 GAACAGATATATTCACACAAACACATCTTGTCCAGGCTGTGCGTGAATGTTACA 969

Qy 374 LeuAspAspLeuAlaAspPheLysGlnGlyValValLysMetAsnLysAspArgThr 393

Db 970 ATTGAGTCTGTGCTATTAATCTTCAAGATGTTGATTAATTAAGCAACAAAGAAACG 1029

Qy 394 GlyGlnProMetGlnLysLysLysTyrLeuAspLysGlyLysThrGlyLysProLysGlyAspAla 413

Db 1030 GAGCAGCCCATGATTAATTTGTAACAAGAGGAATCGCAAGCTGAAGAGAGGCA 1089

Qy 414 ThrValSerTyrGlnAspProProThrAlaLysAlaAlaValGluTrpPheAspGlyLys 433

Db 1090 ACGGTCTCTTTGATGATACCACTTCAGCTTAAGCAGCTAATGACGTGTTGATGTAA 1149

Qy 434 AspPheGlnGlySerLysLeuLysValSerLeuAlaArgLysLysProProMetAsnSer 453

Db 1150 GAATTCCTCCGAATCTCTTCAAGGTCTCATTTGCTACTGCGGGCAGACTTAAAT-- 1206

Qy 454 MetArgGlyLysLeuProProArgGlnGlyArgGlyMetProProProLeuAspArgGly 473

Db 1207 ---CGGGGTGTGGCAATGTGTGAGGCCAGG-----CGAGGAGGA 1248

Qy 474 ProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMetGlyLysArgGlyAsp 493

Db 1249 CCATGCGCGGTGAGAGCTATGAGGTGT-----GGCAGTGTGTGTGTGGC 1296

Qy 494 ArgGlyLysPheProProArgGlyProArgGlySerArgGlyAsnProSerGlyGly 513

Db 1297 CAGAGAGATTTCCACGTGAGGT-----GGTGGCGGT 1329

Qy 514 AsnValGlnHisArgAlaAspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsn 533

Db 1330 GAGAGCAGACAGAGAGCTGTGATGAGAGTCCATATCCACCTGTGAGATATGAAC 1389

Qy 534 PheAlaTrpArgThrGlnCysAsnGlnCysLysAlaProLysProGlnGlyPheLeuPro 553

Db 1390 TTCTCTTGAGAGAAATGAATGACCAAGTGTAAAGCCCTTAAACCAAGATGAC----- 1440



```

QY 554 ProProhepProProGlyYAspArgGlyYArgGlyYProGlyYMetArg--- 572
DB 1441 -----CCAGAA-----GGGGACCAAGTGGCTCTCACATG 1470
QY 573 GlyYArgGlyYLeuMetAspArgGlyYProGlyYMetPheArgGly----- 590
DB 1471 GGGGGTAACCTACAGGGATGATCGTCGTGTGGCAGAGAGGCTATATGACGGGGTAC 1530
QY 591 ---GlyYArgGlyYAspArgGlyYLeuPheArgGlyYArg---GlyMetAspArgGly 608
DB 1531 CGGGGGCCGGCGGGGACCTGTGAGGCTTCCAGGGGGCCGGGTGTGGGACAGAGGT 1590
QY 609 GlyPheGlyYArgYArgArgGlyYProGlyYProProGlyYProLeuMetGluGln 628
DB 1591 GGGCTT----- 1596
QY 629 MetGlyYArgArgGlyYArgGlyYProGlyYLeuMetAsp---LysGlyYGluHis 647
DB 1597 -----GGCCTTGGCAAGATGATTCACAGGGGTGAGCAC 1629
QY 648 ArgGlnGluArgArgAspArgProTyr 656
DB 1630 AGACAGGATCGCAGGAGGAGGCGGTAT 1656

```

## RESULT 9

```

US-10-755-889-649
/ Sequence 649, Application US/10755889
/ Publication No. US20040171823A1
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
/ FILE REFERENCE: D0284 NP
/ CURRENT APPLICATION NUMBER: US/10/755, 889
/ PRIOR FILING DATE: 2004-01-13
/ PRIOR APPLICATION NUMBER: U.S. 60/440, 068
/ PRIOR FILING DATE: 2003-01-14
/ PRIOR APPLICATION NUMBER: U.S. 60/469, 757
/ NUMBER OF SEQ ID NOS: 823
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 649
/ LENGTH: 1822
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-755-889-649

```

## Alignment Scores:

```

Pred. No.: 1 65e-83 Length: 1822
Score: 1281.50 Matches: 299
Percent Similarity: 53.59% Conservative: 59
Best Local Similarity: 44.76% Mismatches: 155
Query Match: 35.27% Indels: 155
Gaps: 29

```

US-10-791-017a-2 (1-656) x US-10-755-889-649 (1-1822)

```

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaIaGlnGlnGlyTyrSerAla 20
DB 73 ATGGCTCAACGATTTATACC-----CAACAACCAACCAAGCTATGGGGCC 120
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAla---GlnThrThrGlnAlaTyrGlyGln 39
DB 121 TACCCACACCGAGCCGGGCGGCTATTCACACAGAGCAGTCCGCTACGAGCAGAG 180
QY 40 SerTyrGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrAla 59
DB 181 AGTTACAGTGTATACCGCTCAGCGAC-----ACTTCA 216
QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProThrGlnGlyTyrThr 79
DB 217 GGCTATGGCCAGAGCAGCTAT---TCTTCTTATGGCCAGAGCCAGAAAGCTATGAACT 273

```

```

QY 80 ProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrGlyAlaTyr 98
DB 274 CAGTCAACTCC-----CAGGATATGGCTCGACCTGGGCTAT 312
QY 99 AspThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyr 118
DB 313 GGC-----AGTACCAAGAGCTCCAACTCTTAC 342
QY 119 GlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArg 138
DB 343 GGGCAGAGAGCTCTATCTGCTATGGCCAGACAGCAGCT----- 384
QY 139 ProGlnAspGlyValnysProThrGlnThrSerGlnProGlnSerThrGlyY Tyr 158
DB 385 -----CCACGACGACCTCGGAAGTTAC 408
QY 159 AsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySer 178
DB 409 GGTAGCAGTTCT-----CAGACGACGACTTATGGCAGCCCCAG---AGTGGAGC 456
QY 179 TyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThr 198
DB 457 TACAGCCAGAGCCT-----AGCTATGTGGA--- 483
QY 199 GlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSer 218
DB 484 -----CAGCAGCAAGCTATGACACAGCAAGCTATATATCCCTCCAG 528
QY 219 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThr 238
DB 529 GGCTATGACAGCAAGCAAGCTATCAACAGCAGAGTGTGTGAGTGTGAGTGTGAGGT 588
QY 239 SerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSer 258
DB 589 GGA-----GGTACTATGGCCAGATCAATCTCCATGATGATGGTGT 633
QY 259 SerSerTyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyr 278
DB 634 GGCAGTGTGTGGCTTATGCAATCAAGACAGAGTGTGAGTGTGAGTGTGAGTGTAT 693
QY 279 GlyGlnGlnSerGlyYLeuPheSerGlyProGlyGluAsnArgSerMetSerGlyProAsp 298
DB 694 GACACAGCAGACCGGTGA-----GGCCGGCGCAGGGGTGTGCACTGCT---GCC 738
QY 299 AsnArgGlyYArgGlyYArgGlyYLeuPheAspArg-----GlyYLeuSerArgGlyY 316
DB 739 GCGCGCGCGCGCGCGGTGTGTGTTCAACCGCAGCAGTGTGTGCTATGAACCAAGGT 798
QY 317 ArgGlyYLeuYArgGlyYLeuMetGlySerAlaGly-----GluArgGlyYLeuPheAsn 334
DB 799 CGTGGAGTGTGGCGGTGTGAGGACAGAGTGTGCGATGGCGGAAGTACCGTGTGCTTCAAT 858
QY 335 LysProGlyYLeuProMetAspGlnGlyProAspLeuAspLeuGlyProProValAspPro 354
DB 859 AATTTGTGGCCCTCGGACCAAGATCACTCATGAC-----TCCGAA 903
QY 355 AspGluAspSerAspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeu 374
DB 904 CAGGATATATGACAAACAAACACCATCTTGTGCAGAGGCTGTGTGAGATATGTTACAT 963
QY 375 AspAspLeuAlaAspPhePheLeuGlnCySGlyValValLeuMetLeuValAspThrGly 394
DB 964 GAGTGTGTGCTGATTAATCTCAAGCAGATTGTGATTATTAAGCAAAAGAAAGAGCGGA 1023
QY 395 GlnProMetIleThrIleTyrLeuAspLysGluThrGlyYLeuProLysGlyAspAlaThr 414
DB 1024 CAGCCATATATTAATTTGACACAGCAGGAAACTGGCACTGTAAGGAGGCAAG 1083
QY 415 ValSerTyrGluAspProProThrAlaValAlaAlaValGluThrPheAspGlyYAsp 434
DB 1084 GTCTCTTTGATGATCACTTCAAGCTTAAAGCAGTATGACTGTTGATGTGAAGAA 1143
QY 435 PheGlnGlySerTyrLeuLysValSerLeuAlaArgYLeuProProMetLeuSerMet 454

```





Qy 201 ThSerTyraPpInserSerTyrsSerGlnGlnAnthrTyrglyGlnProSerSerTy 220  
Db 660 ACTAGNTACATCAGACAGACTTACTCTCAGACAGAACCTTAGGGCCGAGAGACTAT 719  
Qy 221 GlyGlnInserSerTyrglyGlnInserSerTyrglyGln 234  
Db 720 GACAGCAGAGTAGCTATGTGTCAACAAAGCAGCTATGGGACG 761

RESULT 12  
US-10-198-846-9847  
; Sequence 9847, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Little, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steilmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9847  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-9847

Alignment Scores:  
Pred. No.: 3,67e-76 Length: 2299  
Score: 1181.00 Matches: 222  
Percent Similarity: 66.07% Conservative: 0  
Best Local Similarity: 66.07% Mismatches: 1  
Query Match: 32.51% Indels: 114  
DB: 14 Gaps: 1

US-10-791-017a-2 (1-656) x US-10-198-846-9847 (1-2299)

Qy 338 GlyProMetAspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAsp 357  
Db 1072 GGACCCATGATGAGACAGATCTGATCTAGGCCACCTGTAGATCCAGATGAAGAC 1131  
Qy 358 SerAspAsnSerAlaIleTyrglyGlnGlyLeuAsnAspSerValThrLeuAspAspLeu 377  
Db 1132 TCTGACAAAGTGCAATTATGTATCAAGATTAATGACAGTGTGATCTAGATGATCTG 1191  
Qy 378 AlaAspPhePheGlyGlnGlyValAllylMetLeuValGlyGlnGlnProMet 397  
Db 1192 GCAGACTCTTAAAGCTGTGGGTGTGTAAAGTGAAGCAAGAACCTGGGCAACCATG 1251  
Qy 398 IleHisIleTyrlleuAspGlyThrGlyLeuAspGlyAspAlaThrValSerTy 417  
Db 1252 ATCCACATCTTACCTGCGACAGAAAGAACGCCCAAGCCAGTATGCCAGATGCTTAT 1311  
Qy 418 GluAspProProThrAlaValAlaValAluTrpPheAspGlyLeuAspPheGlnGly 437  
Db 1312 GAAAGCCCACTGCGCAAGGCTGCGGTGAATGTTATGGGAAGATTTTCAAGGG 1371  
Qy 438 SerIleLeuLeuValSerIleAlaArglyLeuProMetLeuSerMetAlaGlyGly 457  
Db 1372 AGCAAACTTAAAGCTCTCTGCTCGGAAGAAAGCTCAATGAACATATGCGGGGTGT 1431  
Qy 458 LeuProProArgGlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyPro 477  
Db 1432 CTGCGACCCCGTGGGAGGAGGAGGATGCCACACACATCTCGGTGAGGTCCAGAGGGCCA 1491  
Qy 478 GlyGlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPhe 497

Db 1492 GAGAGCTCTGGGGGACCCATGGGTGCGAGGAGGCGGTGAGAGGATAGAGAGCTTC 1551  
Qy 498 ProProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyValGlnHis 517  
Db 1552 CCTCCAAAGAGAGCCCGGGGTTCCTCCAGAGGAACCCCTCTGAGAGAGAAAGCTCCAGCAC 1611  
Qy 518 ArgAlaGlyAspTrpGlnCysProAsnProGly- 528  
Db 1612 CGAGCTGGAGACTGGAGAGTGTCCCAATCCGTA-TGTACTTGTCTTGGCAATTGATACC 1670  
Qy 528 ----- 528  
Db 1671 TAGAGTGAAGCAACCTTCCTCACCACCATCCCACTAGAGTGATGCTCTGTCTA 1730  
Qy 528 ----- 528  
Db 1731 GAGAAAGAAATGATGACCTGTATGGCTGTGAGACATACTAGCATTGACCTGAGAC 1790  
Qy 528 ----- 528  
Db 1791 GCTTGAAGCTTCTGAAGATTGATTGACCTGTCTGTGGTGAATGCTGCTGAGGC 1850  
Qy 528 ----- 528  
Db 1851 TGTGCCCTAAAGCATGGGTGTACATAGATCTTGTATGTAGTGTGTACTGTTACA 1910  
Qy 528 ----- 528  
Db 1911 CACCACTTCTCTGTTATCTCTCTTGAATGATGATGATGATGATGATGATGAT 1970  
Qy 529 -----CysGlyAsnGlnAsnPheAlaTrpArgThrGlyCysAsnGlnCysGly 544  
Db 1971 GRATGAGGGGTGTGTGAACAGAACTTCGCTGAGAAAGAGAGTCAACAGTGTAA 2030  
Qy 545 AlaProLeuProGlyGlyPheLeuProProProPheProProGly 560  
Db 2031 GCCCAAGCCTGAAGGCTTCTCCGCCACACCTTCCGCCCGGGGT 2078

RESULT 13  
US-09-918-995-27690  
; Sequence 27690, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27690  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(550)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27690

Alignment Scores:  
Pred. No.: 7.63e-61 Length: 550  
Score: 961.00 Matches: 166  
Percent Similarity: 97.65% Conservative: 0  
Best Local Similarity: 97.65% Mismatches: 4  
Query Match: 26.45% Indels: 0  
DB: 10 Gaps: 0

US-10-791-017a-2 (1-656) x US-09-918-995-27690 (1-550)

QY 437 GYSerIySeuIySvAlSerIeulAlArIyLySeProPheCanserMetArgGly 456  
DB 40 GGAGAGCAAACTTNAAGCTCTCTGCTCGAGAAAGCTTCAATGAACAGATATGCGGTGT 99  
QY 457 GYLeuProProArgGluGlyArgGlyMetProProProleuArgGlyGlyProGlyGly 476  
DB 100 GGTCTGCGACCCCTGAGGGGCAAGGCAATGCCACCACTCTCGTAGGTTCAAGAGGC 159  
QY 477 ProGlyGlyProGlyGlyPrometGlyArgMetGlyGlyArgGlyGlyYAspArgGlyGly 496  
DB 160 CCAAGAGGTCTCTGGGGGACCAATGGGTGCGATGGAGAGCCGTGGAGGAGATTAAGAGAGC 219  
QY 497 PheProProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGlyYAsnValGln 516  
DB 220 TTCTCTTCAAGAGAGACCCCGGGGTTCCTCGAGGGAACTCTGAGAGAGAAAGTTCAG 279  
QY 517 HIAArgAlaGlyAspTrpGlnCyProAsnProGlyCyArgIyAsnGlnAsnPhaAlaTrp 536  
DB 280 CACCGAGCTGAGACTGGAGTGTCCCAATCCGGGTGTGAAACAGAACTTCTGCTGG 339  
QY 537 ArgThrGlyCyAsnGlnCySlyAlaProLyAspProGluGlyPheLeuProProPhe 556  
DB 340 AGAAGAGAGTCAACAGGTGAAGGCCCAAGCTGAAGCTTCTCCCGCACCTTT 359  
QY 557 ProProProGlyGlyYAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGly 576  
DB 400 CCGCCCGCGGTGTGTCTGTGCGAGAGTGGCCCTGTGTGCATGCGAGAGAGAGGT 459  
QY 577 GYLeuMetAspArgGlyGlyProGlyGlyMetPheArgGlyGlyYAsnArg 596  
DB 460 GGCTCATGATGATGTGTGTCCCGGTGAATGTTCAATAGTGTGCTGTGTGAGACAGA 519  
QY 597 GYLYGlyPheArgGlyGlyArgGlyMetAsp 606  
DB 520 GGTGGCTTCGTGTGTGCGCGGGGATGAC 549

## RESULT 14

US-09-918-995-9557  
/ Sequence 9557, Application US/09918995  
/ Publication No. US20030073623A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Hyseq, Inc.  
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
/ FILE REFERENCE: 20411-756  
/ CURRENT APPLICATION NUMBER: US/09/918,995  
/ CURRENT FILING DATE: 2001-07-30  
/ PRIOR APPLICATION NUMBER: US/09/235,076  
/ PRIOR FILING DATE: 1999-01-20  
/ NUMBER OF SEQ ID NOS: 38054  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 9557  
/ LENGTH: 568  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (1)...(568)  
/ OTHER INFORMATION: n = A, T, C or G  
US-09-918-995-9557

## Alignment Scores:

Pred. No.: 3,74e-54 Length: 568  
Score: 868.50 Matches: 166  
Percent Similarity: 97.08% Conservative: 0  
Best Local Similarity: 97.08% Mismatches: 4  
Query Match: 23.91% Indels: 2  
DB: 10 Gaps: 1

US-10-791-017A-2 (1-656) x US-09-918-995-9557 (1-568)

QY 307 PheAspArgGlyMetSerArgGlyArgGlyGlyYAsnArgGlyGlyMetGlySer 326

DB 56 TTGATCTGAGAGCATAGACAGAGGTGGCGGAGAGAGACGCGGTGAATGGGC--- 112  
QY 327 AlAGlyGlyuArgGlyGlyPheAsnLyProGlyGlyProMetAspGluGlyProAspLeu 346  
DB 113 GCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGTGAGACCATGATGAAGACCAAGATCTT 172  
QY 347 AspLeuGlyProProValAspProAspGluAspSerAspAsnSerAlaIleYrValGln 366  
DB 173 GATCTAGGCGCCACTGTAGATTCAGATGAAGACTCTGACCAACAGTGCATTTATGTACAA 232  
QY 367 GYLeuAsnAspSerValThrLeuAspLeuAlaAspPhehlyGlnCySlyVal 386  
DB 233 GATTAATATACAGTGTACTTGAATGATCTGACAGACTCTTTAAGCAGTGTGGGTT 292  
QY 387 ValLyMetAsnLyAspThrGlyGlnProMetIleIleIleYrLeuAspLyGluThr 406  
DB 293 GTTAAGATGAACAAGAACTGGGCAACCATATCCATCTGACCAAGAAACA 352  
QY 407 GYLYAspProLyGlyAspAlaThrValSerTyGluAspProProThrAlaYsAlaAla 426  
DB 353 GGAAGGCCCAAGGCCATGCGCACAGTGTCTTATGAAGACCCCACTGCGAAGGCTGCC 412  
QY 427 ValGluTrpPheAspGlyYAspPheGlnGlySerIySeuIySvAlSerIeulAlArG 446  
DB 413 GTGAAATGTTATGTGAGAAATTTTCAGAGGAGCAAACTTAAGTCTCTGTGCTGG 472  
QY 447 LyLeuProProMetAsnSerMetArgGlyGlyLeuProProArgGluGlyArgGlyMet 466  
DB 473 AAGAAGCTTCAATGAACAGTATGCG- GGTGGCTGCGACCCGTAGAGGCAAGAGCATG 531  
QY 467 ProProProleuArgGlyGlyProGlyGlyPro 477  
DB 532 CCAACCACTTTCGTGAGGTCCAGAGAGGCCCC 564

## RESULT 15

US-09-864-761-8001  
/ Sequence 8001, Application US/09864761  
/ Patent No. US20020048763A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Penn, Sharon G.  
/ APPLICANT: Rank, David R.  
/ APPLICANT: Hanzel, David K.  
/ APPLICANT: Chen, Wensheng  
/ TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
/ FILE REFERENCE: Aeomica-X-1  
/ CURRENT APPLICATION NUMBER: US/09/864,761  
/ CURRENT FILING DATE: 2001-05-23  
/ PRIOR APPLICATION NUMBER: US 60/180,312  
/ PRIOR FILING DATE: 2000-02-04  
/ PRIOR APPLICATION NUMBER: US 60/207,456  
/ PRIOR FILING DATE: 2000-05-26  
/ PRIOR APPLICATION NUMBER: US 09/632,366  
/ PRIOR FILING DATE: 2000-08-03  
/ PRIOR APPLICATION NUMBER: GB 24263,6  
/ PRIOR FILING DATE: 2000-10-04  
/ PRIOR APPLICATION NUMBER: US 60/236,359  
/ PRIOR FILING DATE: 2000-09-27  
/ PRIOR APPLICATION NUMBER: PCT/US01/00666  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00667  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00664  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00669  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00665  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00668  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00663  
/ PRIOR FILING DATE: 2001-01-30

```

PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 8001
LENGTH: 540
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC018774.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1,1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1,1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
US-09-864-761-8001

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 21:44:36, Search time 4920 Seconds  
(without alignments)  
5075.236 Million cell updates/sec

Title: US-10-791-017A-2  
Perfect score: 3633  
Sequence: 1 MASTDYSTYQAAAGQCYA.....GGPKNDKGRHRRDRPY 656

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 60479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlp  
-Q/cgml\_1/USPTO.spool.p/US10791017/runat.17022005.125807.22086/app.query.fasta\_1.1358  
-DB=EST -QPM=fastap -SUFFIX=p2n.rst -MIMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=human40.cdd -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=us10791017 @CGML\_1.1.6628 @runat.17022005.125807.22086 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: gb\_eest1.\*  
2: gb\_eest2.\*  
3: gb\_hic.\*  
4: gb\_eest3.\*  
5: gb\_eest4.\*  
6: gb\_eest5.\*  
7: gb\_eest6.\*  
8: gb\_gsest1.\*  
9: gb\_gsest2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2157	3	CR604952 full-leng
2	3633	100.0	2314	3	CR608375 full-leng
3	3618.5	99.6	2159	3	CR625247 full-leng
4	3615	99.5	2103	3	CR619493 full-leng
5	3600.5	99.1	2070	3	CR602561 full-leng
6	3600.5	99.1	2093	3	CR594868 full-leng
7	3600.5	99.1	2126	3	CR610888 full-leng
8	3564.5	98.1	2107	3	AK019460 Mus muscu
9	3549	97.7	2073	3	CR593724 full-leng

10	3549	97.7	2083	3	CR608666 full-leng
11	3530.5	97.2	2373	3	AK034755 Mus muscu
12	3455.5	95.1	2236	3	CR604639 full-leng
13	3321	91.4	2269	3	AK049743 Mus muscu
14	2568	70.7	2692	3	BC043904 Xenopus 1
15	1874	51.6	1379	3	AK014366 Mus muscu
16	1675.5	46.1	4729	3	HS0804802 Homo sapi
17	1672	46.0	1049	5	BX325713
18	1663	45.8	1045	5	BM909438
19	1656	45.6	1008	5	BX353313
20	1645.5	45.3	1058	4	BM802924
21	1629.5	44.9	1056	5	BQ059127
22	1629	44.8	1020	5	BM927718
23	1628	44.8	1015	5	BU841107
24	1628	44.8	1079	5	BX449774
25	1624.5	44.7	1062	5	BM909485
26	1624	44.7	962	5	BX443095
27	1622	44.6	1021	5	BX464988
28	1614	44.4	984	5	BX443496
29	1609.5	44.3	1040	5	BX399070
30	1608.5	44.3	1787	3	CR618632 full-leng
31	1607	44.2	1027	5	BX428312
32	1607	44.2	1062	4	BM473356
33	1607	44.2	1068	5	BX382091
34	1604.5	44.2	1024	4	BM556024
35	1604	44.2	1024	5	BQ062138
36	1602	44.1	993	5	BX462635
37	1602	44.1	1087	5	BX398765
38	1601	44.1	1028	4	BM476095
39	1599	44.0	922	5	BX465100
40	1597	44.0	1013	5	BX367687
41	1589	43.7	1024	4	BM468711
42	1583	43.6	1031	4	BM477969
43	1574.5	43.3	1041	5	BX420692
44	1573	43.3	971	5	BX407826
45	1568	43.2	917	5	BX407419

## ALIGNMENTS

RESULT 1  
CR604952  
LOCUS CR604952 2157 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS01010712 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR604952  
VERSION CR604952.1 GI:50485759  
KEYWORDS HTC; CNSLT cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2157)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2157)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source  
1..2157  
/organism="Homo sapiens"





Db	1961	AAAAAGATTAAAGGACACCCTTATGAGAGCGCAGATTCGGCCCTTAC	
RESULT 2			
LOCUS	CR608375	2314 bp	mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CSODI039YJ24 of Placenta Cot 25-normalized		
ACCESSION	CR608375		
VERSION	CR608375.1	GI:50489182	
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2314)		
JOURNAL	Lt,W.B., Gruber,C., Jesssee,J. and Polayes,D.		
REMARK	Full-length cDNA libraries and normalization		
	Unpublished		
	Contact : Feng Jiang Email : fjiang@life tech.com URL :		
	http://fulllength.invitrogen.com/ invitrogen Corporation 1600		
REFERENCE	Paraday Avenue		
AUTHORS	2 (bases 1 to 2314)		
TITLE	Genoscope.		
JOURNAL	Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr)		
	1st strand cDNA was primed with a NotI-oli5o(4T) primer. Five primes		
	end enriched, double-strand cDNA was digested with Not I and cloned		
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library		
	was normalized. Library was constructed by Life Technologies, a		
	division of Invitrogen.		
FEATRES	Location/Qualifiers		
source	1..2314		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSODI039YJ24"		
	/issue_type="Placenta Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,33e-212	Length:	2314
Score:	3633.00	Matches:	656
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatch:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0
US-10-791-017A-2 (1-656) x CR608375 (1-2314)			
Oy	1	MetAlaserrThrAspTySerThrTySerglnAlaalaaglnglTYrSerAla	20
Db	41	ATGGGTCACGAGATTACAGTAACCTATAGCCAAAGTCGACGCGCACGGGTACAGTGCT	100
Oy	21	TyrThraagiInProthglnGlyTyrAlaaglnThrhglnAlaTYrgylnglnSer	40
Db	101	TACACCGGCCACGCCTCAACAAGATATCACAGACACCAAGGACATAGGGCAACAAGC	160
Oy	41	TyrGlYthrTYrgylnglnProThrsrpValserTYrThrhglnAlaaglnThrhAlaThr	60
Db	161	TATGGAACTTATGAGACACCCACTCATATTCAGCTATACCAAGCTCAGAACCACTGCACC	220
Oy	61	TyrglYglnThrrAlaTYrAlaThrseryTYrgylnglnProProthhglnTYrTh-Th-Pro	80
Db	221	TATGGGACAGACCGCTATGCAACTTCTATGAGACAGCTCCACATGGTTATATACTCCA	280
Oy	81	ThralAProglAlaTYrSerglnProValglnGlyTYrgylThrhglnAlaTYrAspThr	100
Db	281	ACTGCCCCCAGGACATACAGCGAGCTCTCCAGGGGATATGCACTGGTGCTTATGATACC	340

OY	101	ThrThrAlaThrValLTmrThrThrThiAlaSerTyrValAlaIleInserAlaTyrGlyThr	120
Db	341	ACCACGTGACAGTACCACCAACCAGCCCTCTATGAGCTGACTGCAATATAGGCACT	400
OY	121	GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln	140
Db	401	CAGCTGCTTATCCAGCTTATGGGCACAGCCACGACCACTGCACCTTACAAAGCCGAG	460
OY	141	AspGlyAsnLysProThiGlnThrSerGlnProGlnInserSerThiGlyTyrAsnGln	160
Db	461	GATGAAACAAAGCCCACTGAGACTATGCAACTCAATCTAGCACAGGGGGTTTAAACAG	520
OY	161	ProSerLeuGlyTyrTyrGlyInserSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	521	CCAGCCCTAGAGATGTGACACAGATACCTACAGTTATCCCAAGTACCTGGAGGTACCC	580
OY	181	MetGlnProValThrAlaProProSerTyrProProThiSerTyrSerSerThiGlnPro	200
Db	581	ATGACGACAGTCACTGCACCTCCATCTCACTCTACAGCTATTTCTCTACACAGCG	640
OY	201	ThiSerTyrAspGlnInserSerTyrSerGlnGlnAsnThiTyrglyGlnInserSerSerTyr	220
Db	641	ACTAGTATATGATCAGAGCAGTTACTCTCAGAGAAACCTATGGGCAACGACAGCTAT	700
OY	221	GlyGlnGlnInserSerTyrTyrGlyGlnInserSerTyrGlyGlnGlnInserProProThiSerTyr	240
Db	701	GGACACGACAGATACCTATGTGTCAACAAAGCGATATGGCGACGCTCCACATAGTTAC	760
OY	241	ProProGlnThrGlySerTyrSerSerGlnAlaProSerGlnTyrSerGlnInserSerSer	260
Db	761	CCACCCCAACTGATCTTACAGCCCAAGCTCAATATATACCAACAGACAGCAGC	820
OY	261	TyrGlyGlnGlnInserSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln	280
Db	821	TACGGGGACGACAGATTCAATCCGACAGAACCAACCCAGTAGCATGGGGTTATGGGAG	880
OY	281	GluSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerMetSerGlyProAspAsnArg	300
Db	881	GAGCTCTGAGGATTTTCCGGACCGAGAGAGAAACCGGACATGATGGCCCTGATTAACCG	940
OY	301	GlyArgGlyArgGlyGlyPheAspAspArgGlyMetSerArgGlyGlyArgGlyGlyGly	320
Db	941	GGCAGGGAGAGGGGGAGTTTGATCGTGGAGGCATGACAGAGGTGGCGGGAGAGAGA	1000
OY	321	ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet	340
Db	1001	CGCGGTGGAAATGGGCAAGCGCTGGAGACGAGATGGCTTCAATTAAGCTGTGTGACCATG	1060
OY	341	AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn	360
Db	1061	GATGAAAGACACAGATCTTGTATCTTGGCCCACTCTTAATCCAGATGAAGACTCTGACAAC	1120
OY	361	SerAlaIleTyrValGlnGlyLeuAsnAspSerAlaThrLeuAspAspLeuAlaAspPhe	380
Db	1121	AGTCAATTTATATGACAAAGATTAAAGACAGTGTGACTATAGATGATCTGGCAGACTTC	1180
OY	381	PheLysGlnCysGlyValValLysMetAsnLysArgThiGlyGlnProMetIleHisIle	400
Db	1181	TTTAAAGCAGTGTGGGGTCTTTAAGATGAACAAAGAACTGGGCAACCAATCATCAATC	1240
OY	401	TyrLeuAspLysGlnThiGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro	420
Db	1241	TACCTGGAACAAGAAACAGAAAGCCCAAGAGCGATCCACAGTGTCTTATGAAGACCA	1300
OY	421	ProThrAlaLysAlaAlaValGlnThrPheAspGlyLysAspPheGlnGlnLysSerLysLeu	440
Db	1301	CCCACTGCAAGGCTGCGCGGGAATGTATGATGGGAAAGATTTTCAAGGAGCAAACTT	1360
OY	441	LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro	460
Db	1361	AAATCTCTCCCTGCTCGGAAGAAAGCTCTCAATGAACAGTATCGGGGTGTGTCTCCACCC	1420
OY	461	ArgGlnGlyArgGlyMetProProProLeuAspGlyGlyLysProGlyGlyProGlyPro	480

Db	1421	CGTAGGGCAAGGATGCCACCACTCCGTGGAGTCCAGAGGCGCCAGAGTCT	1480
Qy	481	GLYGLYPMeMcGLyArMeMcGLyLYArGLyGLyAbpArGLyGLyPheProProArG	500
Db	1481	GGGGAGCCCAAGGGTCGATGGAGGCGGTGAGAGAGTAAGAGAGGCTTCCCTCCCAAGA	1540
Qy	501	GLYProArGLySArGLyYAbnProSerGLyGLyYaaNValGLNH1ArYGLaGLy	520
Db	1541	GGACCCCGGGGTTCCCGAGGAAACCCCTCTGGAGAGAAAGTCACAGACCGAGCTGA	1600
Qy	521	AsPTpGLyCySProAenProGLyCYSGLYaSnGLnaenPheALATpArGThGLyCyS	540
Db	1601	GACTCGCGAGTCTCCCATCCCGGTTGTGAAACCGAALCTTCGCTCGAGAAACAGAGTGC	1660
Qy	541	AsnGLnCybLYbALaProLYbProGLnGLyPheLupProProProPhProProProGLy	560
Db	1661	AACCAAGETAAAGGCCCCCAAGCCCTGAAAGCTTCCCTCCGCAACCTTTCGCCCCGGGT	1720
Qy	561	GLYAbpArGLyYArGLyGLyYProGLyGLyYMeArGLyGLyYArGLyGLyYLeuMeArSp	580
Db	1721	GGTAGTGTGGCAGAGGTGGCCCTGGTGGATCGGGGAGAGAGGTGGCTCATGGAT	1780
Qy	581	ArGLyGLyYProGLyGLyYMePheArGLyGLyYArGLyGLyYAbpArGLyGLyYPheArG	600
Db	1781	CGTGTGTCTCCGGGGAATGTTCAGAGTGGCCCTGTGTGAGACAGAGGTGGCTTCGGT	1840
Qy	601	GLYGLYArAGLYMeArbArGLyGLyYPheGLyGLyYArArAGLYGLYProGLyGLyY	620
Db	1841	GGTGGCGGGGCATGGACCGAGGTGGCTTTGGTGGAGAAACAGAGTGGCGCTCGGGGG	1900
Qy	621	ProProGLyProLeuMeMcGLnGLnMeMcGLyYArArGLyGLyYArGLyGLyYProGLy	640
Db	1901	CCCCCTGCACCTTTGATGAGAACAGATGGAGAGAAAGAGAGGACCTGGAGGACCTTGA	1960
Qy	641	LYMeArbArLYbGLyNH1ArYGLnGLnArGLaArbArbArbProProTYr	656
Db	1961	AAATATGATTAAGGCGAGCACCGTCAGAGGCGCAAGATCGGCTTAC	2008

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES
CR625247		full-length cDNA clone CSDB005YF1 of Placenta of Homo sapiens (human).	CR625247	CR625247.1	GI:50506054	HTC; CNSLT cDNA.	Homo sapiens (human)	Unpublished				Unpublished					
CR625247		2159 bp mRNA linear HTC 21-JUL-2004						contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/				contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/					
		Placenta of Homo sapiens						Faraday Avenue				Faraday Avenue					
								Genoscope.				Genoscope.					
								2 (bases 1 to 2159)				2 (bases 1 to 2159)					
								Direct Submission				Direct Submission					
								Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :				Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :					
								BP 191 91006 Evry cedex - FRANCE (E-mail : secrete@genoscope.cns.fr				BP 191 91006 Evry cedex - FRANCE (E-mail : secrete@genoscope.cns.fr					
								Web : www.genoscope.cns.fr)				Web : www.genoscope.cns.fr)					
								1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers				1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers					
								end enriched. double-strand cDNA was digested with Not I and cloned				end enriched. double-strand cDNA was digested with Not I and cloned					
								into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library				into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library					
								was normalized. Library was constructed by Life Technologies, a				was normalized. Library was constructed by Life Technologies, a					
								division of Invitrogen.				division of Invitrogen.					
								Location/Qualifiers				Location/Qualifiers					
								1. 2159				1. 2159					
								/organism="Homo sapiens"				/organism="Homo sapiens"					
								/mol_type="mRNA"				/mol_type="mRNA"					

```

/db_xref="taxon:9606"
/clone="CS0DE005YF11"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

```

Alignment Scores:	
Pred. No.:	9.5e-212
Score:	3618.50
Percent Similarity:	99.85%
Best local Similarity:	99.85%
Query Match:	99.60%
DB:	
	3
	Gaps: 1

QY	1	MetLaseRThraPPIYrSerSerThrTySerglnAlaAlaAlaGlnGlnGlnTyYrSerAla	20
Db	46	ATGGGTCACGAGATTACAGTACTATAGCCAAAGCTGACGCGCAGCGGGCTTACAGTCT	105
QY	21	TyThraLagInProthGlnGlnTyYrAlaGlnThrThGlnAlaTyTgLYgInGInSer	40
Db	106	TACACCGCCAGCCCACTCAAGATATATGACAGACCAACCGAGCATATGGCAACAAAGC	165
QY	41	TyTgLYThrTyTgLYgInProThThraPValSerTyTThGlnAlaGlnThrThraLaThr	60
Db	166	TATGGAACCTATGGACAGCCCACTGATCTCAGCTATACCCAGGCTCAGACCACTGAAAC	225
QY	61	TyTgLYgInThraLaTyYrAlaThSerSerTyTgLYgInProProThGlyTyYrThrThPro	80
Db	226	TATGGGAGAGACGGCTATGCAACTCTTATGGACAGCCTCCACATGGTTATACTACTCA	285
QY	81	ThraLaProGlnAlaTyYrSerSerInProValGlnGlyTyTgLYThGlyAlaTyYrAspThr	100
Db	286	ACTCGCCCCCAGGAGATACAGCAGCGCTGTCCAGGGGTATGGACATGGTCTTATATGATAC	345
QY	101	ThrThraLaThraValThrThrThGlnAlaSerTyYrAlaAlaGlnSerAlaTyYrGlyThr	120
Db	346	ACCACTGCTACAGTACACCAACCAACCGGCTCTTATGCAAGCTCAAGTCTGCAATGGCACT	405
QY	121	GlnProAlaTyYrProAlaTyTgLYgInGlnProAlaAlaThraLaProThraPProGln	140
Db	406	CAGCTGCTTATCCAGCCTATGGGACAGCAGCAGCAGCACTGACACTTCAGAACCGCAG	465
QY	141	AspGlyAsnTyYrProThGlnThSerSerInProGlnInSerSerThGlyGlyYrYrGlnGln	160
Db	466	GATGGAACCAAGCCCACTGAGACTATGTAACCTCAATCATTAGCACAGGGGGTTACAACCGAG	525
QY	161	ProSerLeuGlyTyTgLYgInSerAsnTyYrSerTyYrProGlnValProGlySerTyYrPro	180
Db	526	CCCAAGCCTAGGATATGGACAGAGTACATCAAGTTATCCCAAGGTACTCTGGAGCTACCCC	585
QY	181	MetGlnProValThraLaProProSerTyYrProProThSerTyYrSerSerThrGlnPro	200
Db	586	ATGCAAGCAGTACACTGCACTCCATCCATCACTCTTACACAGTATATCTCTACACAGCGG	645
QY	201	ThrSerTyYrAspGlnSerSerSerTyYrSerGlnGlnAsnThrTyYrGlyGlnProSerSerTyYr	220
Db	646	ACTGATTATGATCAGACAGATTACTCTCAGCAAGAACACTTATGGGCAACCGAGCACTAT	705
QY	221	GlyGlnGlnSerSerTyYrGlyGlnGlnSerSerTyYrGlyGlnGlnProProThSerTyYr	240
Db	706	GGAACAGCAGATAGTACTATGTCACAAAGCAGCTATGGGACAGCACTCTCCACTAGTTAC	765
QY	241	ProProGlnThrThGlySerTyYrSerGlnAlaProSerGlnTyYrSerGlnGlnSerSer	260
Db	766	CCACCCCAACTGATCTCTACAGCCAACTCCAACTCAATATATAGCCAAAGACAGCAGAC	825
QY	261	TyTgLYgInGlnSerSerPheArGlnAspHisProSerSerMetGlyValTyTgLYgIn	280
Db	826	TACGGGACGACAGATTCATTCTCGACAGGACCAACCCCAATAGCATGGGTCTTTATATGGCAG	885
QY	281	GlnSerGlyTyYrPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg	300

Db	886	GAGCTGGAGGATTTTCCGACCAAGGAGAAACCGAGCATGAGTGGCCCTGATTAACCG	945
Qy	301	GIYARGLIYARGLIYGLYPHEAPARGLIYGLYMETSERARGLIYGLIYARGLIYGLIY	320
Db	946	GGCAGGGGAMAGGGGGGATTTGATCGTGAAGGCAGACAGAGTGCGGGGAGAGAGA	1005
Qy	321	ARGLIYGLYMETGLYSERIALGLIYGLIYARGLIYGLIYPHEANLVPARGLIYGLIYPROMET	340
Db	1006	CGCGGTGGAAATGGGCG--GCTGGAGAGGAGGTGGCTTCATTAATACCTGGTGGACCCATG	1063
Qy	341	ASPGIUGLIYPROAPLEUAPLEUGLIYPROPROVALAPPROAPGLIYAPSERAPBAN	360
Db	1063	GATBAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTTGACAAAC	1122
Qy	361	SERIALIETRYVALINGLIYLEUANAAPSERVALIIRLEUAPAPLEUANAAPHE	380
Db	1123	AGTCAATTTATGTATCAAGATTAAATGACAGTGTGACTATGATGATCTGGACAGACTTC	1182
Qy	381	PHELYSGINCYREGIYVALIYVALIYMETANLVIYAAATHIRGLIYGLINPROMETLIETHILLE	400
Db	1183	TTTAAGCAGTGTGGGTGTGTTAAGATGAAACAAGAACTGGGCAACCATGATTCACATC	1244
Qy	401	TYRLEUAPLPSGLIUTHIYLYAPPROBPSGLIYAPBALATHVALISERTYGLIYAPPRO	420
Db	1243	THACTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTCTCTATGAMAAACCA	1302
Qy	421	PROTHRALIYVALIALAVALIYUTRPHENAPGLIYVAPAPHEGLINGLIYSERIYLEU	440
Db	1303	CCCACTGGCCAAAGGCTGCCGTGGAAATGGTTTGAATGGAAAGATTTTCAAGGGAGCAAACTT	1362
Qy	441	LYSVALISERLEUAIAYGLIYLYBPROBOMELANISERMETARGLIYGLIYLEUPROPRO	460
Db	1363	AAAGTCTTCCCTTGCTCGGAMAAAGCTTCAATGAAACAGTATGCCGGGTGCTGCCACCC	1422
Qy	461	ARGIUGLIYARGLIYMETPROBROBLEUARGLIYGLIYPROGLIYGLIYPROGLIYGLIYPRO	480
Db	1423	CTGAGGGCAGAGGCATGCCACACACTCGTGGAGGTCCAGAGGGCCCAAGAGTCTT	1482
Qy	481	GLIYGLIYPROMETGLIYARGMETGLIYGLIYARGLIYGLIYAPARGGLIYGLIYPHEPROBPAR	500
Db	1483	GGGGAGCCCAATGGGTCCGATGGAGGCCGTGGAGAGATAGAGAGGCTTCCCTCCACGA	1544
Qy	501	GLIYPROAPARGIYSERARGIYASNPROSERGLIYGLIYAEUNVALIYHIAARGHIAGLIY	520
Db	1543	GGACCCCGGGGTTCCCGAGGGAACCCCTCTGGAGGAGAAACGTCCACACACCGAGCTGGA	1602
Qy	521	ASPTRPGLINCYBPROBANPROGLIYCYDELIVANGIYANENPHEALATTPARGTHIGLIYB	540
Db	1603	GACTGGCAGTGTCCCAATCCGGGTGTGTGAAACCAAGAACTTCGCTGGAGAAACAGAGTGC	1662
Qy	541	ASNGIYNCYBVALIAPROLYBPROGLIYGLIYPHELEUPROPROBOPHEPROBPROBGLIY	560
Db	1663	AACCAAGTGTAAAGGCCCCCAAGCTGTAAGGCTTCTTCCCGCAACCTTTCCGCCCCCGGAT	1722
Qy	561	GLIYAPAPARGIYARGLIYGLIYPROGLIYGLIYMETARGLIYGLIYARGLIYGLIYLEUMETASP	580
Db	1723	GGTATCGTGGCAGAGGTGGCCCTGTGTGCAATGCCGGAGGAAGAGTGGCTCATGAT	1782
Qy	581	ARGGLIYGLIYPROGLIYGLIYMETPHEARGLIYGLIYARGLIYGLIYAPPARGLIYGLIYPHEARG	600
Db	1783	CGTGTGTGTCCCGGTGGAAATGTTCAGAGGTGGCCGTGTGTGAGACAGAGTGTGGCTTCGT	1842
Qy	601	GLIYGLIYARGLIYMETAPAPARGIYGLIYPHEGLIYGLIYATAPAPAGGLIYGLIYPROGLIYGLIY	620
Db	1843	GGTGGCCGGGGCAGTGAACCAAGTGGCTTTGTGTGAGGAAGACAGAGTGGCCCTGGGGGGG	1902
Qy	621	PROBPROGLIYPROLEUMETGLIYGLIYMETGLIYGLIYARGARGGLIYGLIYARGGLIYGLIYPROGLIY	640
Db	1903	CCCCCTGGACCTTTGATGAGAAACAGTGGAGGAAGAAAGAGGAGCGTGGAGGACCTTGA	1962
Qy	641	LYSMETAPPLVSGLIYGLIYHIAARGLIYGLIYATGATGAPAPARGPROTYR	656

DB	1963	AAATGATAAAGCGACACCCTCAGAGGCAGAGATCGGCCCTTAC	2010
RESULT 4			
LOCUS	CR619493	2103 bp	mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CSOD1076Y021 of Placenta Cot 25-normalized		
ACCESSION	CR619493.1	GI:50500300	
VERSION	CR619493.1		
KEYWORDS	HTC; CNSLT_cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 2103)		
JOURNAL	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.		
REMARK	Full-length cDNA libraries and normalized		
	Unpublished		
	Contact : Feng Liang Email : fliang@life.techn.com URL :		
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600		
REFERENCE	Faraday Avenue		
AUTHORS	2 (bases 1 to 2103)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :		
	Bp 191 91006 EVRI cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)		
COMMENT	- Web : www.genoscope.cns.fr		
	1st strand cDNA was primed with a NotI-oligo(4nt) primer. Five primase		
	end enriched, double-strand cDNA was digested with Not I and cloned		
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library		
	was normalized. Library was constructed by Life Technologies, a		
	division of invitrogen.		
FEATURES	Location/Qualifiers		
source	1..2103		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSOD1076Y021"		
	/tissue type="Placenta Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.51e-211	Length:	2103
Score:	3615.00	Matches:	652
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.50%	Indels:	0
DB:	3	Gaps:	0
US-10-791-017A-2 (1-656) x CR619493 (1-2103)			
Oy	5	AspyrserThrrYrsersAlAlAlalagInglYrserAlatYThrAlaIn	24
Dd	1	GATTACGTACTCTTAAGCAAGCGAGCGAGGAGGCTACAGGCTTACACCGCCAG	60
Oy	25	ProthrngInglYrAlaInThThrglnAlatYrGlvgIngsrTYrGlYThrYr	44
Dd	61	CCCACTAGAAGATATGCACAGACCAACGGCATATGGCAACAAGTTAGAACCTAT	120
Oy	45	GLyInPrOthraapValserYThrThrglnAlaInThThralatThrYrGlYgInThr	64
Dd	121	GGAAGGCCCACTAGTCAGCTATATACCAAGGCTCAGACCACTGCACCTATGGGAGACC	180
Oy	65	AlAtYrAlatThrseryTrGlYgInPrObProthrglYrThrThrProthAlaProglIn	84
Dd	181	GCCATATCAACTCTTATATGACACACCTCCCACTGTTATACTCTCCAACTGCCCCAG	240
Oy	85	AlAtYrserglInProvalGInglYrYrGlYThrglYalaYrAspThrThThralatThr	104
Dd	241	GCAATACACCGACGCTGTCCAGGGATATGGCACTGTGCTTATGATATCACCACTGTCTCA	300
Oy	105	VAltThrThrThrglnAlaserYrAlaAlaInserAlatYrGlYThrgInProAlatYr	124

```

Db      ||| 301 GTACCAACCAACCAAGGCTCTTATGAGCTCAGTCTGATATGGCACTGACCGCTTAT 360
Qy      ||| 125 ProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGlnAapGlyAenLys 144
Db      ||| 361 CCAAGCCATGAGGAGGAGCGAGCGAGCGACCTGCACTCAAGACCGAGGATGGAAACAG 420
Qy      ||| 145 ProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAenGlnProSerLeuGly 164
Db      ||| 421 CCACTAGAGCTAGTCACTCACTCACTAGACAGAGGGGTTTACACAGCCCAAGCTTGA 480
Qy      ||| 165 TyrGlyGlnSerAenTyrSerTyrProGlnValProGlySerTyrPrometGlnProVal 184
Db      ||| 481 TATGACAGAGTAACTACAGTTATCCCAAGTACCTGAGAGCTACCCCATGACGACGATC 540
Qy      ||| 185 ThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThrSerTyrAap 204
Db      ||| 541 ACTGCACTCATCTCTCACTCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 600
Qy      ||| 205 GlnSerSerTyrSerGlnGlnAenThrTyrGlyGlnProSerSerTyrGlyGlnGlnSer 224
Db      ||| 601 CAGAGCACTTACTTCAAGAGACACCTTATGGGCAACCGAGCACTTATGACAGCAAGT 660
Qy      ||| 225 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProThrSerTyrProGlnThr 244
Db      ||| 661 AGCTATGCTCAACAAAGCAGCTATGGGAGAGCGCTCCCACTAGTTACCAACCCCAACT 720
Qy      ||| 245 GlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerTyrGlyGlnGln 264
Db      ||| 721 GGAATCTTACAGCCAAAGCTCCAAAGTCAATATAGCCAAAGAGAGAGAGCTACGGGAG 780
Qy      ||| 265 SerSerPheArgGlnAapPhePheSerSerMetGlyValTyrGlyGlnGlnSerGly 284
Db      ||| 781 AGTTCACTTCGACAGAGACACCCAGATGACATGGGTTTATGGGAGAGAGTCTGAGGA 840
Qy      ||| 285 PheSerGlyProGlyGlnAenArgSerMetSerGlyProAapAenArgGlyAargGly 304
Db      ||| 841 TTTTCCGAGACAGAGAGAAACCGAGCATGATGGCTCTGATTAACCGGAGGAGAAAG 900
Qy      ||| 305 GlyGlyPheAapArgGlyGlyMetSerArgGlyGlyValTyrGlyGlnGlnGlyMet 324
Db      ||| 901 GGGGAGATTTATCTGAGAGCATGAGAGAGTGGGAGGAGAGAGAGAGAGAGAGATG 960
Qy      ||| 325 GlySerAlaGlyGlnAargGlyGlyPheAenLysProGlyGlyPrometAapGlnGlyPro 344
Db      ||| 961 GCGAGCCCTGAGAGAGAGAGTGGCTTCAATAGCTGTGTGAGCCCAAGATGAAAGCA 1020
Qy      ||| 345 AapLeuAapLeuGlyProProValAapProAapGlnAapSerAapAenSerAlaIleTyr 364
Db      ||| 1021 GATCTTATCTAGCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Qy      ||| 365 ValGlnGlnLysAenAapSerValThrLeuAapAapLeuAapPhePheLysGlnCys 384
Db      ||| 1081 GTACAAAGATTAATTAATGACAGGTGACTCTAGATGATCTGAGAGCTTCTTAAAGCA 1140
Qy      ||| 385 GlyValValLysMetAenLysArgThrGlyGlnProMetIleHisIleTyrLeuAapLys 404
Db      ||| 1141 GGGGTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1200
Qy      ||| 405 GlnThrGlyLysProLysGlyAapAlaThrValSerTyrGlnAapProProThrAlaLys 424
Db      ||| 1201 GAAACAGAAAGAGCCCAAGGCGATGACAGAGTGCCTTATGAAAGCCCAAGCCCAAG 1260
Qy      ||| 425 AlaAlaValAlaGlnThrPheAapGlyLysAapPheGlnLysLysLysValSerLeu 444
Db      ||| 1261 GCTGCGCTGAAATGTTTGAATGGGAAAGATTTTCAGAGGAGCAAACTTAAAGTCTCC 1320
Qy      ||| 445 AlaArgLysLysProPrometAenSerMetArgGlyGlyLeuProProAapGlnGlyAarg 464
Db      ||| 1321 GCTCGAAGAAAGCTCCAAAGTAAAGATGCGGGGTGTCTGCCAACCCCGTGAAGGAGGA 1380
Qy      ||| 465 GlyMetProProProLeuAargGlyGlyProGlyGlyProGlyGlyProGlyGlyPromet 484

```

---

```

Db      ||| 1381 GGCATCCCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAGTCTCGGAGCAACCAT 1440
Qy      ||| 485 GlyArgMetGlyGlyValArgGlyGlyAapArgGlyGlyPheProProAapGlyProArgGly 504
Db      ||| 1441 GGTGCAATGGAGAGCCGTGAGAGATGAGAGAGGCTTCTCCCAAGAGAGAGAGAGAG 1500
Qy      ||| 505 SerArgLysAapProSerGlyGlyAenValGlnHisAargAlaGlyAapTyrGlnCys 524
Db      ||| 1501 TCCGAGAGGAGAGCCCTCTGAGAGAGAGAAAGTCAACACCGAGCTGAGACTGGAGTGT 1560
Qy      ||| 525 ProAapProGlyCysGlyAenGlnAenPheAlaTyrArgThrGlyCysAenGlnCys 544
Db      ||| 1561 CCAATCCGGGTGTGGAACCAAGCTTCCGTGAGAAACAGAGTGAACCAAGCTGTAA 1620
Qy      ||| 545 AlaProLysProGlnGlyPheLeuProProProPheProProProGlyGlyAapArgGly 564
Db      ||| 1621 GCCCAAGCTGAAAGCTTCTCCGCAACCTTTCGGCCCGGGGTGTATGTTGGC 1680
Qy      ||| 565 ArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAapArgGlyGlyPro 584
Db      ||| 1681 AGAGTGGCCCTGTGTGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy      ||| 585 GlyGlyMetPheArgGlyGlyArgGlyGlyAapArgGlyGlyPheArgGlyGlyAargGly 604
Db      ||| 1741 GGTGAATGTTCAGAGGTGCGCTGTGAGACAGAGGTGGCTTCCGTGTGGCCGCGGCG 1800
Qy      ||| 605 MetAapArgGlyGlyPheGlyGlyValArgAargGlyGlyProGlyGlyProProGlyPro 624
Db      ||| 1801 ATGAGCCAGAGTGTGCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy      ||| 625 LeuMetGlnGlnMetGlyGlyAargArgGlyGlyAargGlyGlyProGlyLysMetAapLys 644
Db      ||| 1861 TTGATGAAACAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy      ||| 645 GlyGlnHisArgGlnGlnAargAargAapArgProTyr 666
Db      ||| 1921 GCGAGACACCTGTCAGAGCGCAGAGATCGGCCCTAC 1956

```

RESULT 5  
CR602561  
LOCUS  
DEFINITION  
Full-length cDNA clone CS0D1068YN3 of Placenta Cot 25-normalized of Homo sapiens (human).  
ACCESSION  
CR602561  
VERSION  
CR602561.1 GI:50483368  
KEYWORDS  
HTC; cNSLT; cDNA.  
SOURCE  
Homo sapiens (human)  
Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 2070)  
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact : Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue  
2 (bases 1 to 2070)  
REFERENCE  
Genoscope.  
Direct Submissions  
Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
Location/Qualifiers  
1..2070  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

/clone="CS0DI068YN23"  
 /cissue="Plicenta Cot 25-normalized"  
 /plasmid="PCWVSF01\_6"

## ALIGNMENT SCORES:

Pred. No.:	1,15e-210	Length:	2070
Score:	3600.50	Matches:	651
Percent Similarity:	99.85%	Conservative:	0
Best Local Similarity:	99.85%	Mismatches:	0
Query Match:	99.11%	Indels:	1
	3	Gaps:	1

US-10-791-017a-2 (1-656) x CR602561 (1-2070)

QY 5 AspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAlaTyrThrAlaGln 24  
 Db 1 GATTACAGTACCTTATGACCAAGCTGACGCGCAGGCGCTTACAGTGTACACGCCAG 60

QY 25 ProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSerTyrGlyThrTyr 44  
 Db 61 CCACCTCAAGGATATGACAGACCAACCGGCTATGGGCAACCACTATGGAACCTAT 120

QY 45 GlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlyGlnThr 64  
 Db 121 GGACAGCCCACTGATGCTATGACCTATACAGGCTGACACCACTGACCTATGGCAGACC 180

QY 65 AlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrProThrAlaProGln 84  
 Db 181 GCGTATGCACTTATGAGAGCGCTCCACCTGTTATACTTCAACCTGCCCCCAG 240

QY 85 AlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThrThrThrAlaThr 104  
 Db 241 GCATACAGCAGCGCTGTCAGGGGTATGCGACTGCTGTTATGATACACCACTGCTACA 300

QY 105 ValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThrGlnProAlaTyr 124  
 Db 301 GTCCACCAACCCAGGCTCTCTATGACCTGATGCTGATATGACCTGACCTGCTTAT 360

QY 125 ProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGlnAspGlyAsnLys 144  
 Db 361 CCAGCCTATGGGAGCAGCCAGCAGCAGCAGCTTACCAAGACCGGAGTGAACAAG 420

QY 145 ProThrGlnThrSerGlnProGlnSerSerThrGlyTyrGlnGlnProSerLeuGly 164  
 Db 421 CCACCTGAGACTGCTCAACTCAATCTATGACAGGGGTATCAACACCGCCAGCTGAGA 480

QY 165 TyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrProMetGlnProVal 184  
 Db 481 TATGACAGAGTATCACTGATTAATCCCAAGTACCTTGAAGCTACCCATGACAGCTGC 540

QY 185 ThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThrSerTyrAsp 204  
 Db 541 ACTGACACTCATCTCTACCTCTCAACAGGATATTCCTTACAGCGGAGCTAGTTAGAT 600

QY 205 GlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyrGlyGlnGlnSer 224  
 Db 601 CACAGCAGTATCTCTGACAGAACCACTATGGGCAACCGAGCACTATGACAGCAAGT 660

QY 225 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyrProProGlnThr 244  
 Db 661 AGCTATGCTCAACCAAGCAGCTATGGGAGAGCTCCCACTGATTAACCAACCCCAACT 720

QY 245 GlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSerTyrGlyGlnGln 264  
 Db 721 GGATCTCAACGCAAGCTCCAGTCAATATAGCAACAGAGCACTACGGGAGAG 780

QY 265 SerSerPheArgGlnAspAlaProSerSerMetGlyValTyrGlyGlnGlnSerGly 284  
 Db 781 AGTTCAATTCGAGCAGCAACCCCAATGACATGGGTATATGGGAGAGTCTGAGAGA 840

QY 285 PheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArgGlyTyrGlyArg 304

Db 841 TTTTCCGACCAAGAGAGAACCCGAGCATAGTGCCTCTATTAACCGGGGACAGGAGAGA 900

QY 305 GlyGlyPheAspArgGlyGlyMetSerArgGlyGlyYArgGlyGlyYArgGlyGlyMet 324  
 Db 901 GGGGATTTTATGCTGAGGAGCATGACAGAGTGGGCGGAGGAGGAGCGCGGTGATG 960

QY 325 GlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMetAspGlnGlyPro 344  
 Db 961 GGC--GCTGAGAGCCAGGTGCTTCAATTAAGCTGCTGAGACCAATGATGAAGAGCA 1017

QY 345 AspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsnSerAlaIleTyr 364  
 Db 1018 GATCTTATCTTAGGCCCACTGATGATCCAGATGAAGACTCTGACCAAGTCAATTTAT 1077

QY 365 ValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPheLeuGlnCys 384  
 Db 1078 GTACAAAGATTAAATGACAGTGTGACTCTGATGATCTGACAGCTCTTTAAGCAATGT 1137

QY 385 GlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIleTyrLeuAspLys 404  
 Db 1138 GGGGTGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCACATCTACCTGGACAG 1197

QY 405 GlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysAspProProThrAlaLys 424  
 Db 1198 GAAACAGAAAGCCCAAGGCGATGCCACAGTCTCTATGAAGACCAACCACTGCCAG 1257

QY 425 AlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeuValSerLeu 444  
 Db 1258 GCTGCCGTGAAATGTTTGAATGGAAAGATTTTCAAGGAGAGCAACTTAAGTCTCCCTT 1317

QY 445 AlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProProAlaGlyLysArg 464  
 Db 1318 GCTCGAAGAAAGCCTCCAAATGAACAGTATCGGGGTGCTGACCCCTGAGGGAGA 1377

QY 465 GlyMetProProProLeuAspArgGlyGlyProGlyGlyProGlyGlyProMet 484  
 Db 1378 GGCAATGCCACCACTCCTGATGATCCAGAGGCCCAAGAGTCTGGGGAGCCATG 1437

QY 485 GlyArgMetGlyYArgGlyGlyYAspArgGlyGlyPheProProAlaGlyProArgGly 504  
 Db 1438 GGTGCGATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCAAGAGACCCCGGGGT 1497

QY 505 SerArgGlyAsnProSerGlyGlyYAsnValGlnHisArgAlaGlyAspTyrGlnCys 524  
 Db 1498 TCCGAGGAGAACCCCTCTGAGAGAGAAAGTCCAGACCTGAGACTGGCACTGT 1557

QY 525 ProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrPargThrGluCysAsnGlnCysLys 544  
 Db 1558 CCGAATCCGGGTGTGGAACCAAGACTTGCCTGAGAAACAGAGTCCACAGTGTAG 1617

QY 545 AlaProLysProGlnGlyPheLeuProProProPheProProProGlyGlyYAspArgGly 564  
 Db 1618 GCCCAAGGCTGAAGGCTTCTCCGCCACCTTCCGCCCGGGGTGTGATCGTGGC 1677

QY 565 ArgGlyGlyProGlyGlyMetLargGlyGlyYArgGlyGlyYLeuMetAspArgGlyPro 584  
 Db 1678 AGAGGTGGCCCTGTGTCATGCGGAGAGAAAGTGGCTCTATGATCTGTGTGTGCC 1737

QY 585 GlyGlyMetPheArgGlyGlyYArgGlyGlyYAspArgGlyGlyPheArgGlyGlyYArgGly 604  
 Db 1738 GGTGGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGTGCTTCCGTGTGGCCGGGCG 1797

QY 605 MetAspArgGlyGlyPheGlyGlyGlyYArgArgGlyGlyProGlyGlyProProGlyPro 624  
 Db 1798 ATGACCGAGAGTGGCTTTGTGTGAGAGAAAGAGTGGCCCTGGGGGGCCCTCGAGCT 1857

QY 625 LeuMetGlnGlnMetGlyGlyYArgArgGlyGlyYArgGlyGlyProGlyGlyYMetAspLys 644  
 Db 1858 TTGATGAAACAGATGGAGAGAAAGAGAGAGCTGGAGAGACTGGAAAAATGAGATAA 1917

QY 645 GlyGlnHisArgGlnGluArgArgAspArgProTyr 656  
 Db 1918 GCGGAGCACCTGACAGAGCGCAGAGATCGGCCCTTAC 1953





QY 485 G1yArGMeTc1yG1yArGg1yG1yAspArGg1yG1yPheProProArGg1yProArGg1y 504  
 Db 1438 GGTGCAATGGAGAGCCCTGAGAGATGAGAGGCTTCCCTCCMAAGAGACCCCGGGGT 1497  
 QY 505 SerArg1yAnPProSerG1yG1yG1yAnValG1nH1sArG1aG1yAspTTPG1nCy 524  
 Db 1498 TCCCGAGGAGACCCCTTGGAGAGAGACGTCCAGACCGAGCTGAGACTGGCACTGT 1557  
 QY 525 ProAnPProG1yCyGg1yAnG1nAnPheAlaTPrArG1yG1yCyAsAnG1nCy9Lys 544  
 Db 1558 CCCAATCCGGGTGTGGAAACCAAACTTCGCTCGAGAAACAGAGTCACCAAGTGAAG 1617  
 QY 545 AlaProLyBProG1yG1yPheLeuProProProPheProProG1yG1yAspArGg1y 564  
 Db 1618 GCCCAAGCCCTGAGAGCTTCTCCGCCACCTTTCGCCCGGGGTGTGATCGTGGC 1677  
 QY 565 ArgG1yG1yProG1yG1yMeTArGg1yG1yArGg1yG1yLeuMeTAspArGg1yG1yPro 584  
 Db 1678 AGAGGTGGCCCTGTGGCAATGGGAGAGAGAGAGAGTGGCTCATGATCGTGTGGTCCC 1737  
 QY 585 G1yG1yMeTArGg1yG1yArGg1yG1yAspArGg1yG1yPheArGg1yG1yArGg1y 604  
 Db 1738 GGTGGAATGTTCAAGAGTGGCCGCGTGTGAGAGACAGAGTGGCTTCGTTGGTGGCGGGC 1797  
 QY 605 MetAspArGg1yG1yPheG1yG1yG1yArGg1yG1yProG1yG1yProProG1yPro 624  
 Db 1798 ATGAGCCGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTCGCGGGGCCCTCGAGACT 1857  
 QY 625 LeuMeTc1yG1yMeTc1yG1yArGArGg1yG1yArGg1yG1yProG1yG1yPheMeTAspLy 644  
 Db 1858 TTATGGAACAGATGGAGAGAGAGAGAGAGAGTGGAGACTGGAAATATGATAAA 1917  
 QY 645 G1yG1nH1sArG1nG1nArGArGAspArGProTy 656  
 Db 1918 GCGAGACCCGTCAAGAGAGCGAGATCGGCCCTAC 1953

RESULT 7  
 CR610888 2126 bp mRNA linear HTC 21-JUL-2004  
 LOCUS  
 DEFINITION  
 (human).  
 full-length cDNA clone CS0DF037Y107 of Fetal brain of Homo sapiens

ACCESSION CR610888  
 VERSION CR610888.1 GI:50491695  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2126)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 CONTACT: Peng Liang Email: filiang@life.rockefeller.edu  
 URL: http://fulllength.invitrogen.com/InvitrogenCorporation1600

REFERENCE 2 (bases 1 to 2126)  
 AUTHORS Faraday Avenue  
 TITLE Genoscope.  
 JOURNAL Direct Submission  
 SUBMITTED (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
 Web : www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.

FEATURES  
 Location/Qualifiers  
 1..2126  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF037Y107"

ORIGIN /tissue\_type="Fetal brain"  
 /plasmid="pCMVSPORT\_6"

Alignment Scores:  
 Pred. No.: 1,18e-210 Length: 2126  
 Score: 3600.50 Matches: 651  
 Percent Similarity: 99.85% Conservative: 0  
 Best Local Similarity: 99.85% Mismatches: 0  
 Query Match: 99.11% Indels: 1  
 DB: Gaps: 3

US-10-791-017A-2 (1-656) x CR610888 (1-2126)

QY 5 AspTySerThrTySerGlnAlaAlaGlnGlnG1yTySerAlaTyThrAlaGln 24  
 Db 1 GATTACAGTACCTATAGCCAAAGCTGACGAGCAGACAGAGGCTCAAGTGTTCACCGCCAG 60  
 QY 25 ProThrGlnG1yTyThrAlaGlnThrThrGlnAlaTyG1yGlnGlnSerTyG1yThrTy 44  
 Db 61 CCCACTCAAGGATATGACAGACACCCAGGCAATATGGGCAACAAAGCTATGAACTAT 120  
 QY 45 G1yG1nProThrAspAlaSerTyThrGlnAlaGlnThrThrAlaTyThrTyG1yGlnThr 64  
 Db 121 GAGACGCCACTGATGTCACTATACCCAGGCTCAAGACCTGCAACTATGGGCAAGC 180  
 QY 65 AlaTyAlaThrSerTyG1yGlnProProThrG1yTyThrThrProThrAlaProGln 84  
 Db 181 GCCTATGCAACTTCTATGAGACAGCTCCCACTGATTATCTACTCACTGCCCCAG 240  
 QY 85 AlaTySerGlnProValGlnG1yTyTyThrG1yAlaTyThrAspThrThrThrAlaThr 104  
 Db 241 GCATACAGCCAGCGCTGTCCAGGGGATAGGCACTGGCTTATATATACCACTGCTTACA 300  
 QY 105 ValThrThrThrGlnAlaSerTyThrAlaAlaGlnSerAlaTyG1yTyThrGlnProAlaTy 124  
 Db 301 GTACACACCAACCGAGCTCTCTATGAGAGCTCACTGATGTGATATGAGCTGCTTAT 360  
 QY 125 ProAlaTyG1yGlnGlnProAlaAlaThrAlaProThrArGProGlnAspG1yAsnLy 144  
 Db 361 CCAAGCTATGGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420  
 QY 145 ProThrGlnThrSerGlnProGlnSerTyThrG1yTyThrGlnProSerLeuGly 164  
 Db 421 CCCACTGAGACTATGATCACTCAATCTACACAGGGGGTTACCAACAGCCAGCTTACAGA 480  
 QY 165 TyG1yGlnSerAsnTySerTyProGlnAlaProG1ySerTyProMetGlnProVal 184  
 Db 481 TATGACAGATTAAGTATATCCCAAGTACCTGGAGCTTACCCCATGACGCAATGTC 540  
 QY 185 ThrAlaProProSerTyProProThrSerTySerSerThrGlnProThrSerTyAsp 204  
 Db 541 ACTGACCTTCACCTTACCTCTTACAGTATATCTTACACAGCGAGCTATATGAT 600  
 QY 205 GlnSerSerTySerGlnGlnAlaThrTyG1yGlnProSerSerTyG1yGlnGlnSer 224  
 Db 601 CAGAGCAATTAATCTCAGAGAACACCTATGGGCAACAGAGCTATGACAGCAGAGT 660  
 QY 225 SerTyG1yGlnGlnSerSerTyG1yGlnGlnProProThrSerTyProProGlnThr 244  
 Db 661 AGCTATGTCACAAAGAGAGCTATGGGAGAGCTTCCCATGTTATCCACCCCAACT 720  
 QY 245 G1ySerTySerGlnAlaProSerGlnTySerGlnGlnSerSerSerTyG1yGlnGln 264  
 Db 721 GATCTTACAGCCAACTCAAGTATATATGACCAAGAGCAGCAGCTATCGGAGCAG 780  
 QY 265 SerSerPheArGlnAspHisProSerSerMetG1yAlaTyG1yGlnGlnSerG1yGly 284  
 Db 781 AGTTCAATCCGACAGACCAACCCAGTACATGGGGTATATGGGAGAGTGTGAGGA 840  
 QY 285 PheSerG1yProG1yGlnAsnArGSerMetSerG1yProAspAnArG1yArG1yArG 304  
 Db 841 TTTTCCGAGACAGAGAGAAACCGAGCATATGAGCTTGTATACCGGGGAGGAGAAAG 900

[illegible]

LOCUS	AK019460	2107 bp	mRNA	linear	HTC 03-Apr-2004
DEFINITION	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830417B1 product:Ewing sarcoma homolog, full insert sequence.				
ACCESSION	AK019460				
VERSION	AK019460.1 GI:12859676				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493574				
PUBMED	11042159				
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Kikunishi, T., Tashiro, H., Itoh, M., Suni, M., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yonezaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409, 685-690 (2001)				
MEDLINE	11076861				
PUBMED	11076861				
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
MEDLINE	11076861				
PUBMED	11076861				
AUTHORS	6 (bases 1 to 2107)				
TITLE	Aadachi, J., Aizawa, K., Akahira, S., Akiyama, T., Arai, A., Aono, H., Arawaka, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hisakawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kikunishi, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, K., Saito, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-AUG-2000) Yoshitake Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)				



```

Db 1274 CCAAGTCGAAGGCTGCGTGGAAATGTTTGGATGGAAATTTTCAAGGAGGAACCTT 1333
Qy 441 Lyva1Ser1eua1aaaglygylgylProProMeCaBSeMeCharg1yglYleuProPro 460
Db 1334 AAAGTGTCTTGTGCGCAAAAGAACCTTCCATGAACGCAATGCGGAGGACATGCCACT 1393
Qy 461 ArgGlu1yArgg1yMetProProProleuAarg1yglYProglYglYProglYglYPro 480
Db 1394 CGTGAAGGCGAGGGAGATGCCACCACTTCTGTGAAGTCTGTGTGCGCCAGAGAGCCCT 1453
Qy 481 G1yglYProMeG1yArgMeG1yglYArgg1yglYApAarg1yglYpHeProProArg 500
Db 1454 GGAAGAACCAATGGGTCCGATGGAGGCGCGTGAAGAGACAGAGGGGCGCTTCCCTCAAGA 1513
Qy 501 G1yProAarg1ySerAarg1yAenProSeG1yglYglYAenValG1nH1aG1aagly 520
Db 1514 GGGCGCCGAGGCTCCAGAGAAACCCCTGTGAAGAGAAATGCCAGCACCGAGCTGGA 1573
Qy 521 AspTrog1nCysProAenProG1yCysG1yAand1nAenHea1aTPaGthnGluCys 540
Db 1574 GACTGGCAATGCTCCCATCCGGGCTGTGAACCAAGACTTCCGTTGGAAGACAGAAATGC 1633
Qy 541 AasnG1nCys1yAa1aProLySProglYglYpHeLeuProProPheProProProglY 560
Db 1634 AACCAAGTAAAGGCGCTTAAGCCGAGGCGCTTCCGCGCACCTTCCACCTCCGGGT 1693
Qy 561 G1yAAspAarg1yAarg1yglYProglYglYMeCaArg1yglYArgg1yglYleuMeAap 580
Db 1694 GGTGATGTGAACAGAGTGGCGCTGTGTGCAATGCGAGAGAGAGAGAGAGAGAGAGAG 1753
Qy 581 ArgG1yglYProglYglYMetPheAarg1yglYArgg1yglYApAarg1yglYpHeArg 600
Db 1754 CGTGTGTCTCTGAGAGAAATCTTCAAGGTGCGAGAGTGGAGACAGAGAGAGCTTCCGA 1813
Qy 601 G1yglYArgg1yMeAAspAarg1yglYpHeG1yglYglYArgg1yglYProglYglY 620
Db 1814 GGTGCGCGTGAATGAGACCAAGGTGCTTGTGTGAAGAGAGACAGAGTGTCTCGGGGG 1873
Qy 621 ProProglYProleuMeG1nMerG1yglYArgg1yglYArgg1yglYArgg1yglYProglY 640
Db 1874 CCTCTGTGACCTTAAATGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1933
Qy 641 LyMeAAspLyg1yGlu1aArgG1nGlu1aArgAargAargProLy 656
Db 1934 AAATGATTAAGCGAGACCGTCAAGAGACGAGAGACCGGCGCTTAC 1981

RESULT 9
CR593724 2073 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR593724
VERSION CR593724.1 GI:50474531
KEYWORDS HTC; CNSLT; cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1. (bases 1 to 2073)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue
2 (bases 1 to 2073)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

```

```

end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..2073
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1011Y005"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 1.65e-207 Length: 2073
Score: 3549.00 Matches: 639
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.69% Indels: 0
DB: 3 Gaps: 0
US-10-791-017A-2 (1-656) x CR593724 (1-2073)
Qy 18 TySerAlaTyThrAlaGlnProThnGlnGlyTyAlaGlnThrThnAlaTyGly 37
Db 3 TACAGTCTTACACCCCGAGCCCACTCAAGATATGCAAGACCAACCGAGATATGG 62
Qy 38 GlnGlnSerTyGlyThrTyGlyGlnProThnAAspAserTyThrGlnAlaGlnThr 57
Db 63 CAACAAAGCTATGAACTATGAGACGCCACATGATGATGATGATGATGATGATGATGATG 122
Qy 58 ThrAlaThrTyGlnGlnThrAlaTyAlaThrSerTyGlyGlnProThnGlyTy 77
Db 123 ACTGACCTATGGAGACCGCTATGCACTTCTTATGAGAGCTCCACCTGTTAT 182
Qy 78 ThrThProThAlaProGlnAlaTySerGlnProValGlnGlyTyGlyThnGlyAla 97
Db 183 ACTACTCAATGCGCCCGAGGATACAGCGCTGTCCAGGGATATGCACTGGTCT 242
Qy 98 TyAAspThrThrAlaThrValThrThnGlnAlaSerTyAlaAlaGlnSerAla 117
Db 243 TATGATACCAACCACTCTCACTGACACACACCGAGCTCTTATCAGCTCATGCA 302
Qy 118 TyGlyThnGlnProAlaTyProAlaTyGlyGlnGlnProAlaThrAlaProThr 137
Db 303 TATGCACTGAGCTCTTATCCAGCTATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
Qy 138 ArgProGlnAAspGlyAenLySProThnGlnThnSerGlnProGlnSerThnGly 157
Db 363 AGACCGCAGAGATGAACAAGCCACTGAGACTGACACTCAACTCAACTCAAGGGGT 422
Qy 158 TyAAspGlnProSerLeuGlyTyGlyGlnSerAenTySerTyProGlnValProGly 177
Db 423 TACAAACAGCCCGCTAGATATGAGACAGAGATCAAGTATATCCAGGATACCTGGG 482
Qy 178 SerTyProMeGlnProValThrAlaProProSerTyProProThnSerTySer 197
Db 483 AGTACCCAGTGAAGCAGTCACTGCACTTCACTTCACTTCACTTCACTTCACTTCT 542
Qy 198 ThrGlnProThSerTyAAspGlnSerTySerGlnGlnAenThnTyGlyGlnPro 217
Db 543 ACACAGCCGATAGTATATGATGACAGTACTCTCAGAGAAACCTATGGCAACCG 602
Qy 218 SerSerTyGlyGlnGlnSerSerTyGlyGlnGlnSerSerTyGlyGlnGlnPro 237
Db 603 AGCAGCTATGACAGAGAGTATGATGATCAACAAGAGAGAGAGAGAGAGAGAGAG 662
Qy 238 ThrSerTyProProGlnThnGlySerTySerGlnAlaProSerGlnTySerGln 257
Db 663 ACTAGTACCAACCCCAAGCTGATCTTCAAGCCAGAGCTCAAGTCAATATGACAG 722
Qy 258 SerSerSerTyGlyGlnGlnSerSerPheArgGlnAAspHisProSerSerMetGlyVal 277

```



Db	243	TATGATACCACTGCTCACTGATCAACCAACCGGCTCTTATGCAAGCTCAGTCTGCA	302
Qy	118	TyrGlyYThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr	137
Db	303	TATGGCACTCAGCGCTTATTCAGACCTATGAGGAGAGACGACGACCACTGCACTTCA	362
Qy	138	ArgProGlnaAPGlyValenLysProThrGlyLunThrSerGlnProGlnSerSerThrGlyLys	157
Db	363	AGACCGCAGGATGGAAAGAACCCCACTAGACTACTGCACTCAATCTTACGACAGGGGGT	422
Qy	158	TyrAenGlnProSerLeuGlyTyrGlyGlnSerAenTyrSerTyrProGlnValaProGlyLys	177
Db	423	TACAAACAGCCCAAGCTTAGATATGGAACAGAGTAACTAAGTTATCCCAAGTACTTGG	482
Qy	178	SerTyrProMetGlnProValThrAlaProSerTyrProProThrSerTyrSerSer	197
Db	483	AGTACCCCACTGACAGCAGTCACTGCACTTCACTCACTCACTCACTCACTCACTCACT	542
Qy	198	ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnaenThrTyrGlyGlnPro	217
Db	543	ACACAGCGCACTAGTATGATCAAGGACGTTACTTCGACGAAACCTTATGGCGAACCG	602
Qy	218	SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnPro	237
Db	603	AGCAGCTATGGACAGCAGATGCTATGATGTCACAAAGCAGCTATGGCAGAGCTTCCC	662
Qy	238	ThrSerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGln	257
Db	663	ACTAGTATCCCAACCCCAATCGAATCTTACAGCCAAAGCTTCAATATATGCCAACAG	722
Qy	258	SerSerSerTyrGlyGlnGlnSerSerPheArgGlnaenLysProSerSerMetGlyVal	277
Db	723	AGCAGCAGCTTACGGGCAAGCAGATTCATTCGACAGGACCAACCCCACTAGCAGATGGGTCTT	782
Qy	278	TyrGlyGlnGlnSerGlyGlyPheSerGlyProGlyGlyLysArgSerMetSerGlyPro	297
Db	783	TATGGGCAAGAGTCTGGAGATTTTCCGACCAAGAGAAACCGAGCATGATGAGGCTT	842
Qy	298	AspAenArgGlyArgGlyArgGlyLysPheAspArgGlyGlyMetSerArgGlyLysArg	317
Db	843	GATTAACGGGGCAAGGGGAAAGGGGGATTGATCGTAGAGCATGACGAGGTGGGGG	902
Qy	318	GlyGlyGlyArgGlyGlyMetGlySerAlaGlyLysArgGlyGlyPheAsnLysProGlyLys	337
Db	903	GGAGAGAGACCCGGGTGGATGGGCGCCCTGGAGAGCAGAGGTGGCTTCAATAGACCTGGT	962
Qy	338	GlyProMetAspGlnGlyProAspLysAsnLysGlyLysProValAspProAspGlnLys	357
Db	963	GGACCCATGATGAAAGAGACAGATCTTATCTTAGGCCCACTGTATGATCCAGATGAAC	1022
Qy	358	SerAspAenSerAlaLysTyrValGlnGlnLysLysAenAspSerValThrLysAspAsnLys	377
Db	1023	TCTGACAAACAATGCAATTTATGTCACAGAGTTAATATGACAGTGCATCTTATGATGATCTG	1082
Qy	378	AlaAspPhePheLysGlnCysGlyValLysMetAenLysArgThrGlyGlnProMet	397
Db	1083	GCAACACTCTTTAAGCAGTGTGGGTGTATTAAGTAAGAACAGAACAGGGCAACCCCATG	1142
Qy	398	IleIleIleTyrLysLysAspLysGlnThrGlyLysProLysGlyLysAspAlaThrValSerTyr	417
Db	1143	ATCCACATCTTACCTGGACAAAGAAACAGAAAGCCCAAGGCGAGTCCACAGTGTCTTAT	1202
Qy	418	GlyAspProProThrAlaLysAlaAlaValGlyTyrPheAspGlyLysAspPheGlnGlyLys	437
Db	1203	GAAACACCACTGACCTGCAAGGCTGCTGTAATGTTGATGGAAAGATTTTTCAGAGG	1262
Qy	438	SerLysLysLysValSerLysLysAlaArgLysLysProProMetAsnSerMetArgGlyLys	457
Db	1263	AGCAAACTTAAAGTCTCCCTTGTCTCGAAGAAAGCTTCAATGAACAGATATCGGGGTGT	1322
Qy	458	LeuProProArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyPro	477

Db	1323	CTGCAACCCCGTAGGGGAGAGGAGATGCACACCACTCCGTGGAGGTCCAGAGAGGCCCA	1382
Qy	478	GIYGIYPRoGIYGIYPRoMeGIYARMeGIYGIYARGIYGIYAPARGIYGIYPhE	497
Db	1383	GGAAGTCCCTGGGGAGCCCATGGGTCCGATGGGAGCCGTGGAGAGATAGAGAGGCTTC	1442
Qy	498	PRoPRoARGIYPRoARGIYsEArGIYAPARPRoErgIYGIYAPARValGIHhE	517
Db	1443	CTCCCAAGAGGAGCCCGGGGTTCCCGAGGGAGAACCCCTCTGGAGGAGGAAGGTCACAGAC	1502
Qy	518	ARGIaGIYAPRTFPIYCYsPRoAsnPRoGIYCYsGIYAsnGIYAsnPhEaIaTRpAR	537
Db	1503	CGACTGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAGAACCTCGCTGAGAGA	1562
Qy	538	ThrGIYCyAsnGIYCYsLYsAIIaPRoLYsPRoGIYGIYPhEaUPRoPRoPhEPRo	557
Db	1563	ACAGAAGTCAACCAAGTGAAGGCCCAAGCCTGAAGGCTTCCTCCGCCACCTTCCTCG	1622
Qy	558	PRoPRoGIYGIYAPARGIYARGIYGIYPRoGIYGIYMeIARsGIYGIYARGIYGIY	577
Db	1623	CCCCCGGGTGTGTATCGTGGCAGAGTGGCCCTGTGTGCATGCGGGAGAGAGAGTGGC	1682
Qy	578	LeuMeARsPARGIYGIYPRoGIYGIYMeIARsPhEArGIYGIYARGIYGIYAPARGIY	597
Db	1683	CTCATGTAGATGTGTGTGTCCCGGTGGAAATGTTCAGAGAGTGGCCGTGTGGAGACAGAGT	1742
Qy	598	GIYPhEArGIYGIYARGIYMeIARsPARGIYGIYPhEGIYGIYGIYARGIYGIY	617
Db	1743	GGCTTCGTGGTGGCCGGGGGATGCAGCCAGAGTGGCTTTGGTGGAGAGACGAGTGGC	1802
Qy	618	PRoGIYGIYPRoPRoGIYPRoLeuMeGIYIIMeGIYGIYARGIYGIYARGIYGIY	637
Db	1803	CTGTGGGGGGCCCCCTGGACCTTTGATGTGAACAGATGGAGAGAGAGAGAGAGCTGGA	1862
Qy	638	GIYPRoGIYLYsMeIARsAPLYsGIYGIYHhIARsGIYIuARsARsPARsPRoTYr	656
Db	1863	GGACTCGGAAAAATGATTAAGGCGACACCGTTCAGAGCCGACAGATCGGCCCTTAC	1919
RESULT 11			
LOCUS	AK034755		
DEFINITION	Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:943003IM11 product:Swing sarcoma homolog, full insert sequence.		
ACCESSION	AK034755		
VERSION	AK034755.1	GI:26084174	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Mullaly; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kaishigaki, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.		

**TITLE** RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

**JOURNAL** Genome Res. 10 (11), 1757-1771 (2000)

**MEDLINE** 20530913

**PUBMED** 11076861

**REFERENCE** 4

**AUTHORS** The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

**TITLE** Functional annotation of a full-length mouse cDNA collection

**JOURNAL** Nature 409, 685-690 (2001)

**REFERENCE** 5

**AUTHORS** The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL** Nature 420, 563-573 (2002)

**REFERENCE** 6 (bases 1 to 2373)

**AUTHORS** Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Seno, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission

**JOURNAL** Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT** CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

**FEATURES** Please visit our web site for further details.

**source** URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

**FEATURES** 1. .2373

**source** /organism="Mus musculus"

**FEATURES** /mol\_type="mRNA"

**FEATURES** /strain="C57BL/6J"

**FEATURES** /db\_xref="FANTOM\_DB:9430031M11"

**FEATURES** /db\_xref="taxon:10090"

**FEATURES** /clone="9430031M11"

**FEATURES** /tissue\_type="embryonic body between diaphragm region and neck"

**FEATURES** /clone\_id="RIKEN full-length enriched mouse cDNA library"

**FEATURES** /dev\_stage="12 days embryo"

**FEATURES** 43. .2011

**FEATURES** /note="Ewing sarcoma homolog (MGI:99960, GB|NM\_007968, evidence: BLASTN, 99%, match=2172)

**FEATURES** /putative=""

**FEATURES** 2358. .2363

**FEATURES** /note="putative"

**FEATURES** 2373

**FEATURES** /note="putative"

**ORIGIN**

**Alignment Scores:**

**Pred. No.:** 2,64e-206

**Score:** 3530.50

**Percent Similarity:** 98.48%

**Best Local Similarity:** 97.11%

**Query Match:** 97.18%

**Length:** 2373

**Matches:** 638

**Conservative:** 9

**Mismatches:** 8

**Indels:** 2

DB: 3 Gaps: 1

US-10-791-017a-2 (1-656) x AK034755 (1-2373)

Qy 1 MetAlaSerThrApPThrYrSerThrYrSerGlnAlaAlaGlnGlnGlyYrSerAla 20

Db 43 ATGGCGTCCACGGATTACGATTACGATTACGATTACGATTACGATTACGATTACG 102

Qy 21 TTTThrAlaGlnProThrGlnGlyYrAlaGlnThrGlnAlaYrGlnGlnGln 40

Db 103 TACACGCCGCCAGCACTCAAGATATGACAGACGCCAGCATATGAGGCAACAAG 162

Qy 40 TTTGlyYrThrGlyGlnProThrApPAlaSerYrThrGlnAlaGlnThrAlaTh 60

Db 163 CTATGAACTTATGACAGGCTCTAGTGGTATGAGCTTATGAGCTTATGAGCTTAC 222

Qy 60 TTTGlyYrGlnThrAlaYrAlaYrAlaThSerYrGlyGlnProProThrGlyYrThr 80

Db 223 CTACGGGCAACATGCAATGCACTTCTTACGACGCTTCCACTGTTATGACACTCC 282

Qy 80 oThrAlaProGlnAlaYrSerGlnProValGlnGlyYrGlyYrThrAlaYrApPTh 100

Db 283 AACTGCCCCGACAGCGTACAGCCAGCTGTGAGGATATGAGCTGAGCTTATGACAG 342

Qy 100 TTTThrAlaThValThrThrThrGlnAlaSerYrAlaAlaGlnSerAlaYrGlyTh 120

Db 343 CACCACTGCTACAGTCCACCAACGAGGCTCTTACGAGCTTACGACTCATATGACAC 402

Qy 120 TGTProAlaYrProAlaYrGlyGlnGlnProAlaAlaThAlaProThrApPProG 140

Db 403 CCAAGCTGCTTACCTCCACCTATGAGCCAGCAACGCAACGCACTTACCAACCA 462

Qy 140 nApPAlaYrAlaYrProThrGlnThrThrSerGlnProGlnSerYrThrGlyYrAla 160

Db 463 GATATGTAACAAGCTCTGAGCTGATGATCACTCAATATGACAGGGGTTATACCA 522

Qy 160 nProSerLeuGlyYrGlyGlnSerAlaYrSerYrProGlnValProGlySerYrPr 180

Db 523 ACCAGCTTATGATATGACAGATGACATGACATGACATGACATGACATGACATG 582

Qy 180 oMetGlnProValThrAlaProProSerYrProProThrSerYrSerSerThrGlnPr 200

Db 583 AATGCAACGATACGACGACCTTCACTTCTTCAACGATCTTCTTCAACGAC 642

Qy 200 oThrSerYrApPAlaSerSerYrYrSerGlnAlaThrYrGlyYrAlaProSerSerY 220

Db 643 GACTAGTTACGATCAGGAGGATTAAGTCTCAGCAACACTATGGGAGCGAGCTA 702

Qy 220 TGTGlnGlnSerSerYrGlyGlnGlnSerSerYrGlyGlnGlnProProThrSerY 240

Db 703 TGGACAAACAGATGATATGTCMAACAGACCTATGGGACAGCTCTCTATGTTA 762

Qy 240 rProProGlnThrGlySerYrSerGlnAlaProSerGlnYrThrSerGlnGlnSerSe 260

Db 763 CCCCCCTCAGACTGATCTTACGACGAGCTCCAAATCAATATACCAACAGACAG 822

Qy 260 rYrGlyGlnGlnSerSerYrThrGlnAlaPAlaProSerSerMetGlyYrAlaYrGly 280

Db 823 CTACGGGCAACAGATTAATTCGACAGACCAACCAAGATGAGTGTATATGGCA 882

Qy 280 nGlnSerGlyYrPhaSerGlyProGlyGlnAlaYrSerMetSerGlyYrProApPAla 300

Db 883 GAGGTGAGAGATTTTCGAGACAGAGAAACGAGCTTATGAGGCTCTATTAACG 942

Qy 300 GGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 320

Db 943 GGGCAGGAGAGAGGAGATTTATGATGAGGAGTATGAGGAGTATGAGGAGGAG 1002

Qy 320 YATGATGlyMetGlySerAlaGlyAlaYrGlyYrPhaSerYrProGlyYrProMe 340

Db 1003 ACGCGGTGAGCTGGG---GCTGAGAGAGAGGTGCTTCAATTAACCTGTGAGACCAT 1059

Qy 340 tAepGlnGlyYrProApPAlaYrProProValApPAlaYrProApPAlaYrProApP 360



```

Db 1060 GAGTGAAGACAGATCTTGAATCTAGAGCTTCCTCATATGATCCCATGAAGACTCTGACAA 1119
Qy 360 nseraiailetyrvalnqlyleuaaspservaitheleuaspleualaaeph 380
Db 1120 CAGGCAATTTATCTGACAGATTAAATGACAAATGATCTGATGATCTGACAGACTT 1179
Qy 380 ephelyaglycysglyvalvallylmetclmlylvarxthglyvglnpmetctlehsil 400
Db 1180 CTTTAAAGAGTGGGGTTGTCAAGATGACAAAGAACTGGACACCTGATGACATAT 1239
Qy 400 eYrLeuAspLygluThrGlyLyPProLyglYAspAlaThrValSerTyrgluAspPr 420
Db 1240 CTACCTGATAGAGAGACAGAAAGCCTTAAAGGGGAGCCACAGTGTCTATGAAGATCC 1299
Qy 420 oPProThAlaAlaAlaAlaValGluTTPheAspGlyLyAspPheGlnGlySerlyle 440
Db 1300 ACCAACTGCAAAAGGCTCCCTGGATGGTTGATGGGAAAGATTTTCAGAGAACT 1359
Qy 440 uLyValSerLeuAlaArgLySlyPProPmetAsnSerMetArgLyglYleuPProP 460
Db 1360 TTAAGTCTCTTCCCGAAAGAGCTCCATATGACAGATGGGGAGGCGATGCCACC 1419
Qy 460 oATgLuGlyArgLyMetPProPProLeuArgLyglYProGlyglYProGlyglYPr 480
Db 1420 TCGTGAAGGAGGAGGAGATGACACACCACTTCGTGAGGTCTGTGGCCAGAGAGCCC 1479
Qy 480 oGlyglYProMetGlyArgMetGlyglYArgLyglYAspArgLyglYphePProPAr 500
Db 1480 TGGAGGACCATGGTGGCATGGAGGCGCTGGAAACCAAGACTTGTGGAAACAGATG 1539
Qy 500 gGlyPProArgLySerArgLyAsnProSerGlyglYglYAsnValGlnHisArgAla 520
Db 1540 AGGGCCCCGAGGCTCCAGAGAAACCCCTCTGAGAGAAATGTCAGACAGAGCTGG 1599
Qy 520 yAspTPrgLncysProAsnProGlyCyglYAsnGlnAsnPheAlaTrrpArgThrgLucy 540
Db 1600 AGACTGGCAGTGTCCCAATCAAGGCGCTGGAAACCAAGACTTGTGGAAACAGATG 1659
Qy 540 sAenGlnCyAlaAlaAlaProLyPProGlyglYpheLeuPProPProPhePProPProG 560
Db 1660 CAACCACTGTAAGCCCTTAAGCCCGAGGCTTCTCCCGCACCTTTCACCTCCGGG 1719
Qy 560 yGlyAspArgLyArgLyglYProGlyglYMetArgLyglYArgLyglYleuMetAs 580
Db 1720 TGGTGAATCGTGAAGAGTGGCTGCTGTCATGCGAGAGAGAGAGAGAGACTGAGA 1779
Qy 580 pArgGlyglYProGlyglYMetPheArgLyglYArgLyglYAspArgLyglYphear 600
Db 1780 CCGTGGGTCTCTGAGAGATGTTCAAGGTGGCAGAGTGGAGACAGAGAGGCTTCCG 1839
Qy 600 gGlyglYArgLyMetAspArgLyglYpheGlyglYglYArgArgLyglYProGlyglY 620
Db 1840 AGGTGGCGGTGAATGACAGAGGTGCTTGTGGAGAGAGAGAGAGAGTGTCTTGGGG 1899
Qy 620 yProPProGlyProleuMetGlnMetGlyglYArgArgLyglYArgLyglYProGly 640
Db 1900 GCTTCTCGAACCTTTATGAGACAGATGGAGAGAAAGGCGGAGCTGAGAGACTGG 1959
Qy 640 yLyMetAspLyglYlunHisArgGlnGluArgArgAspArgProTy 656
Db 1960 GAAATATGATTAAGGCGAGCACCTCAGAAACGAGAGACCGGCTTAC 2008

```

```

RESULT 12
LOCUS CR604639 2236 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DJ013YN08 of T cells (Jurkat cell line)
ACCESSION CR604639
VERSION CR604639.1 GI:50485446
KEYWORDS HTC; cDNA;
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2236)
TITLE Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
JOURNAL Full-length cDNA libraries and normalization
Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Paradise Avenue
Genoscope.
2 (bases 1 to 2236)
REFERENCE
AUTHORS Direct Submision
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Source
1..2236
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ013YN08"
/risue_type="T cells (Jurkat cell line) cot
10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 9 74e-202 Length: 2236
Score: 3455.50 Matches: 632
Percent Similarity: 96.34t Conservative: 0
Best Local Similarity: 96.34t Mismatches: 1
Query Match: 95.11t Indels: 24
DB: 3 Gaps: 2
US-10-791-017a-2 (1-656) x CR604639 (1-2236)
Qy 1 MetAlaSerThrAspTySerThrTySerGlnAlaAlaGlnGlnGlyTySerAla 20
Db 19 ATGGCTCCACCGGATTAAGTACTTATGACCAAGCTGACGAGCGGAGGCTTACAGTCT 78
Qy 21 TyTThAlaGlnPProThrglnGlyTyAlaGlnThThGlnAlaTyrglyGlnGlnSer 40
Db 79 TACACCGCCCGAGCCCACTCAAGATATGACACAGACCAAGGATATGGGCAACAAAGC 138
Qy 41 TyrglyThTyrglyGlnPProThAspValSerTyThGlnAlaGlnThThAlaThr 60
Db 139 TATGAACTTATGAGACAGCCCACTATGTCAGTATACCAAGCTCAGACCTGCAAC 198
Qy 61 TyrglyGlnThThAlaTyAlaThrSerTyrglyGlnPProThrglyTyTThThPro 80
Db 199 TATGGCAGACCGCTATGCACTTCTTATGACAGCTCCCACTGATTATACTCCA 258
Qy 81 ThAlaPProGlnAlaTySerGlnPProValGlnGlyTyrglyTyThGlyAlaTyAspThr 100
Db 259 ACTGCCCCCGAGCATAACGAGCGCTGTCAGGGGTATGGCATGTGCTTATGATACC 318
Qy 101 ThThAlaThTyAlaThThThThGlnAlaSerTyAlaAlaGlnSerAlaTyrglyThr 120
Db 319 ACCACTGCTACAGTACCAACCAAGGCTCTCTATGACAGCTCAGTCTGATATGGACT 378
Qy 121 GlnProAlaTyProAlaTyrglyGlnGlnProAlaAlaThThAlaPProThArgProGln 140
Db 379 CAGCGCTTATTCAGACTATGGGAGGAGCAGCAGCAGCAGCAGCTTCAAGACCGCAG 438
Qy 141 AspGlyAsnLyPProThrglnThThSerGlnPProGlnSerSerThThGlyTyTAsnGln 160
Db 439 GATGAAACAAAGCCCACTAGACTATGTAACCTCAATCTATGACACAGGGGTTCAACAG 498

```

QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 499 CCAGGCTGAGATGAGACAGAGTAATCAAGTTATCCAGGATGAGTACTGGAGACTACCC 558  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 559 ATGCACCACTGCTGACCTCCATCTCACTCTCAACGACTTCTTCAACAGCCG 618  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 DB 619 ACTAGTTATGATCAGAGCAGTACTCTCAGCAGAAACCTATGGGACAGCAGCTAT 678  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 DB 679 GGCACACAGATGCTATGCTCAACAAAGCAGTATGGGACAGCAGCTCCACTAGTAC 738  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 DB 739 CCACCCCAACTGGATCTCAACCCAGCTCCAGTCAATATGACCAACAGACAGCAGC 798  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 799 TACGGCAGACAGATTCATTCAGACAGACACCCAGTACATGGGTGTTATGGGACG 858  
 QY 281 GlnSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 859 GAGCTGAGAGATTTTCCGACACAGAGAACCCGAGCATGTAGTCCCTGATTAACCGG 918  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 DB 919 GCGAGGGGAAAGAGGGGATTTGATCGTGAGCATGACGAGAGTGGCCGGAGAGAGA 978  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 DB 979 CCGGGTGAATGGGCGGCTGAGAGCAGAGTGGCTTCAATAGCTGGTGAACCCATG 1035  
 QY 341 AspGlyGlyProAspLysAspLysGlyProProValAspProAspGlyLysAspAsn 360  
 DB 1036 GATGAAGGACCAAGATCTGATCTAGGCCCCAGCTGATCAATGAACTGTGACAC 1095  
 QY 361 SerAlaIleTyrValGlnGlyLysAsnAspSerValThrLysAspLysAlaAspPhe 380  
 DB 1096 AGTGCATTTTATTAACAAGATTAAATGACAGTGTACTTAATGATCTGGCAGACTTC 1155  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1156 TTTAACAGAGTGGGGTGTGTAAGATGAAACAAGAACTGGGCAACCATGATCCATC 1215  
 QY 401 TyrLysAspLysGlyLysThrGlyLysProLysGlyLysAspAlaThrValSerTyrGlyLysAspPro 420  
 DB 1216 TACCTGACAAAGAAACAAGAAAGCCCAAGGAGTCCACAGTGTCTTATGAAGACCC 1275  
 QY 421 ProThrAlaLysAlaValAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1276 CCACACTGCCAAGGCTGCGTGGATGTTGATGGGAAAGATTTTCAAGGAGCAAACTT 1335  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLysLeuProPro 460  
 DB 1336 AAAGTCTCCCTGCTCGAAGAGAGCCCTCAATGAACGATGCGGGGTGTCTGCCACCC 1395  
 QY 461 ArgGlyGlyArgGlyMetProProProLysAspGlyGlyProGlyGlyProGlyGlyPro 480  
 DB 1396 CCGAGAGGCGAGAGCAGTCCACCACTCCGAGAGGTCCAGAGAGGCCCAAGAGGTCTT 1455  
 QY 481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyLysAspArgGlyGlyPheProProArg 500  
 DB 1456 GGGGAGACCATGGGTCCAGTGGAGGCGCTGGAGAGATGAGAGGCTTCCCTCCAAAG 1515  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyLysValGlnHisArgAlaGly 520  
 DB 1516 GGAACCCCGGGGTTCGAGAGGAAACCTCTGAGAGAGAAACGTCCAGCAGCAGCTGGA 1575  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyLysGlnAsnPheAlaTyrPheThrGlyCys 540

DB 1576 GACTGCGAGTGTCCCAATCCGGGTGTGAAACCAAACTTCCCTGGAGAACAGAGTCC 1635  
 QY 541 AsnGlnCysLeuLysAlaProLysProGlyGlyPheLeuProProProPheProProGly 560  
 DB 1636 AACCACTGTAAAGGCCCAAGCCTGAAGGCTTCTCCGCGAACCTTTCCGGCCCCGGGT 1695  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLysMetAsp 580  
 DB 1696 GGTGATCGTGGCAGAGGTGGCCCTGTGGCATGCGGGGAGAGAGAGTGGCTCCATGAT 1755  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyLysAspArgGlyPheArg 600  
 DB 1756 CGTGTGTCTCCGCTGAATGTTCAGAGGTGCGCTGTGTGAGACAGAGGTGCTTCCTG 1815  
 QY 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 620  
 DB 1816 GGTGGCCGGGGCATGAGCCGAGGTGGCTTTGTGTGAGAGAAAGAGAGTGGCCCTGGGGGG 1875  
 QY 621 ProProGlyProLeuMetGlnMetGlyArgArgGlyGlyArgGlyGlyProGly 640  
 DB 1876 CC-CCA----- 1880  
 QY 641 LysMetAspLysGlyGlyLysArgGlnGlyLysArgArgAspArgProTyr 656  
 DB 1881 -----GGCAGACCGCTCAGAGCGCAGAGATGGCCCTAC 1916  
 RESULT 13  
 AK049743 2269 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK049743  
 DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length  
 full insert sequence.  
 ACCESSION AK049743  
 VERSION AK049743.1 GI:26093614  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Komuro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Komuro, H., Akiyama, J., Nishi, K., Kitsuana, T., Taahito, H., Itoh, M.,  
 Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, O.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20510913  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE 5

## AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**TITLE**

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

**JOURNAL  
REFERENCE  
AUTHORS**

6 (bases 1 to 2269)

TITLE  
JOURN

The FANTOM consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2269)  
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanganbuchi, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsumura, T., Miyazaki, A., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsumi, N., Okazaki, Y., Saito, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shihagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, I., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

**COMMENT**

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>

## FEATURES

```

source
1..2269
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:C530046A18"
/db_xref="taxon:I0090"
/clone="C530046A18"
/tissue_type="spinal cord"
/clone_lib="RIKEN full-length enriched mouse cDNA library
/dev stage="12 days embryo"
48..1903
/note="Xing sarcoma homolog (MGD|MGI:99960, GB|NM_007568
evidence: BLASTN, 99%, match=2172)
putative"
2251..2256
/note="putative"
2269
/note="putative"

```

## ORIGIN

Alignment Scores:	
Pred. No.:	1,74e-193
Score:	3321.00
Percent Similarity:	93.45%
Best Local Similarity:	92.38%
Query Match:	91.41%
DB:	3
Gaps:	1
Length:	2265
Matches:	606
Conservative:	7
Mismatches:	5
Indels:	39

US-10-791-017A-2 (1-656) X AK049743 (1-2269)

Qy	1	METALSERTHIRAPYRSEERTHRYRSGINALAALAGLGINOLYRSEERLA	20
Db	48	ATGGCGTCCACGGAATTACGACTCTATGTCAAGTCGACCACGAGGCCTACAGTCT	10
Qy	21	TyrThAlaGlInProThrgInGlnLYTYrAlaGlnThThGlnAlaTyrcLyngInSer	40
Db	108	TACACGCCCGCAGCACTCAAGATVTGACAGACCAACCCAGCATATGGCGAACAAAGC	16

Qy	41	TyrGlyThrPheTyrGlyInpProThraAspValSerTyrThrGlnAlaGlnThrThraAlaPhe	60
Db	168	TATGAACCTATGAGACAGCTTACTGATGTAAGCTATACAGGCTCAACACCTGGCACG	227
Qy	61	TyrGlyGlnThrAlaTyrAlaIaThrSerTyrGlyGlnProProThnglyTyrThrThraPro	80
Db	228	TACGGGAGAGCTGATATGCAACTTCTTATGGGACAGCTCTCCACTGGTATTATGACCTCCA	287
Qy	81	ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyValaTyrAspThr	100
Db	288	ACTGCCCCGAGGGGTGTCAGAGCCACTGTGACGGGATATGACACTGGGACTTATATGACAGC	347
Qy	101	ThrThrAlaThrValaTmrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
Db	348	ACCACTCTCTACAGTCACCAACAAGGAGCTTCTTACGAGCTCAAGACTGCAATATATGGACCC	407
Qy	121	GlnProAlaTyrTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThraPProGln	140
Db	408	CAGCTGCTTACCCCACTTATGGCAGACAGCAACAGCAACGGCACTTACCAAGACAG	467
Qy	141	AspGlyAsnLysProThrglnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln	160
Db	468	GATGTAAACAAGCTGCTGAGACTAGTCAACCTCAATCTACACAAGGGGGTATTATACCAA	527
Qy	161	ProSerLeuGlyTyrGlyGlnSerAsnThySerTyrProGlnValaProGlySerTyrPro	180
Db	528	CCGAGCCTTAGATATGACACAGATTAAGTACAGCTATCCCAAGGTACTGGGAGCTAACCA	587
Qy	181	MetGlnProValaThrAlaProProSerTyrProProThraSerTyrSerSerThrGlnPro	200
Db	588	ATGAGAGCAAGTACCGCACTCCATCTTATCTCTACAGCTACTCTCTTACAGGCG	647
Qy	201	ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220
Db	648	ACTGATTACGATCAGACAGACTTACTCTCAGAGAAACCTATGGGAGCCGAGCAGCTAT	707
Qy	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnInpProProThraSerTyr	240
Db	708	GGAACAACAAGATACCTATGTGCAACAAGCAGCTATGGGAGCGCCCTCTACTATGTTAC	767
Qy	241	ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer	260
Db	768	CCCCCTCAGACTGATCTCTACGCGCAGGCTTCAAGTCAATTATACCAACAAGACGACGC	827
Qy	261	TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValaTyrGlyGln	280
Db	828	TACGGGACGACGAGATTCAATCCGACAGAACACCCCACTAGTACATGGGTTTATGGCAG	887
Qy	281	GlnSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg	300
Db	888	GAGCTCTGAGGATTTTCCGACCCGAGAGAGAACCGGAGCTTGAGTGGCCCTGATTAACCGG	947
Qy	301	GlyTyrArgGlyTyrArgGlyGlyPheAspAspArgGlyMetSerAspArgGlyValaArgGly	320
Db	948	GGCAGGGAGAGAGGGGGGATTTGATCGTCGGAGGCACTGACAGA-----	989
Qy	321	ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet	340
Db	989	-----	989
Qy	341	AspGlnGlyProAspLeuAspLeuGlyProProValaAspProAspGlyAspSerAspAsn	360
Db	990	-----GATCCCGATGAAAGCTCTGACAAAC	1013
Qy	361	SerAlaLeuTyrValaGlnGlyLeuAsnAspSerValaThrLeuAspAspLeuAlaAspPhe	380
Db	1014	AGTCCAAATTATGTGCAAGATTAAATGACATATGACTCTGGATGATCTGGCAGACTTC	1073
Qy	381	PheLysGlnCysGlyValaValaLysMetAsnLysArgThrGlyGlnProMetIleHisIle	400
Db	1074	TTTAAAGCAGTGTGGCGTTGTCACATGATCAACAAGAGAACTGGACAAACCATATATCATATTC	1133
Qy	401	TyrLeuAspLysGlnThrGlyLysPProLysGlyAspAlaThrValSerTyrGlyLysAspPro	420



QY 139 ProGlnAspGlyValAsnLysProThrGluThrSerGlnProGlnSerThrGlyTyr 158  
 Db 431 CTTCAAGATGGAAGCAAGCCAGAGATACCAATGACCCCAACAGTACAGCACATAT 490  
 QY 159 AsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySer 178  
 Db 491 GCCAGGCTTGGCTTATGCTAATGCAAGAGCTACAGCTACCTCAAGTGTGTCAGC 550  
 QY 179 TyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThr 198  
 Db 551 TATTCAGTGAAGCAAGTGAAGGACCACTTCACTTCACTTCACTTCACTTCACTTCA 601  
 QY 199 GlnProThrSerTyrAspGlnSer----- 206  
 Db 602 CAGCATCAAGCTATGAAGAGAGCTCTATTCACAGCAAGTCTTACACCAAGAGAGC 661  
 QY 207 -----SerTyrSerGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 Db 662 AGCTATGACAGCAAACTTCAATTTCTCAGAGGCTCTTATGACAGCAAGTATAT 721  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyr-----GlyGlnGln 235  
 Db 722 GGTTCACAAAGTGAATATGACCAAGAGTACCAAGCAAGCAAGCAAGCAAGCAAGT 761  
 QY 236 ProProThrSerTyrProProGlnThrGlySerTyr-----SerGlnAlaProSerGlnTyr 254  
 Db 782 CTTCTTCAAGCTTACCACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 841  
 QY 255 SerGlnGlnSerSerSerTyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSer 274  
 Db 842 GGGCAACAGAGCAAGTATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTCTTACAGT 901  
 QY 275 MetGlyValTyrGlyGlnGln-----SerGlyGlyPheSerGlyProGlyLysAsnArgSer 293  
 Db 902 AAAAGTCCCTATGCGCAGAGGCTCTACAGGTTTCTCCAGCTCTGCGCAATATCTGAT 961  
 QY 294 MetSerGlyProAspAsnArgGlyArgGlyArgGlyValPheAspArgGlyGlyMetSer 313  
 Db 962 TCTGAGAGCTCCGATTCACGACAGAGGCTCGGGGCTGTTTGAAGAGTGGCATGAGC 1021  
 QY 314 ArgGlyGlyValArgGlyGlyValArgGlyGlyMetGlySerAlaGlyLysGlyGlyPhe 333  
 Db 1022 CGTGGTGGAGG---GGCTCTCGTGGGCTATGGC---GGTGAAGATTCAGGCTGGCTTC 1075  
 QY 334 AsnLysProGlyValProMetAspGlnGlyProAspLeuAsnLeuGlyPro----- 350  
 Db 1076 AGTAAAGCTGTGATGCTTGGATGATGATGATGATGATGATGATGATGATGATGAT 1135  
 QY 351 -----ProValAspProAspGlnAspSerAspAsnSerAlaLeuTyrValGlnGly 367  
 Db 1136 TTACCATTTGCTTGGACCCAGAGAGAGCAAGAAACAAATCTGTATGTACAGGGA 1195  
 QY 368 LeuAsnAspSerValThrLeuAspAspLeuAlaAspPheLeuGlnCysGlyValVal 387  
 Db 1196 CTAAAGAGATTAATCTGTGAAGAAATGTGATTTCTTCAACACCTGTGTGATGTC 1255  
 QY 388 LysMetAsnLysArgThrGlyGlnProMetIleIleTyrLeuAspLysGlyGlnThrGly 407  
 Db 1256 AGATTAACAGAGGAGAGAGCAAGCAAGCTTGAACCTTTTTCGAGCAAGAGAGAGG 1315  
 QY 408 LysProLysGlyAspAlaThrValSerTyrGlnAspProProThrAlaLysAlaVal 427  
 Db 1316 AAGCCCAAGAGATGAGCACTGTCTTTTGAAGACCACTTCACTTCACTTCACTTCACT 1375  
 QY 428 GluTrpPheAspGlyLysAspPheGlnGlySerLysLeuValSerLeuAlaGlyLys 447  
 Db 1376 GAACCTTGTATGGAGAGATCTTAATGAGAGCAAGTGAAGTGCCTGCTCCCAAA 1435  
 QY 448 LysProProMetAsnSerMetArgGlyGlyLysProProArgGlyGlyGlyMetPro 467  
 Db 1436 AAGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1495  
 QY 468 ProProLeuArgGlyGlyProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMet 487

Db 1496 CCACTCTCCGT-----GGAGCTCCCATGGGTCGATG 1528  
 QY 488 GlyValArgGlyValAspArgGlyGlyPheProProArgGlyProArgGlySerArgGly 507  
 Db 1529 GGTGGCCGAGAGAGAGACCGTGGGAGATTCATGCCAAGAGAGACAGGTGACCAAGAGT 1588  
 QY 508 AsnProSerGlyValGlyAsnValGlnHisArgAlaGlyAspTrpGlnCysProAsnPro 527  
 Db 1589 AGCCCAAGT---AGCGTATATGTCAGCACAGGAGCTGTGATCTGCACTTCCCAACCCG 1645  
 QY 528 GlyCysGlyAsnGlnAsnPheAlaTrpArgThrGlnCysAsnGlnCysLysAlaProLys 547  
 Db 1646 GATGTGAAGAACCAAAATTTTGTCTTGAGAGAACGATATGTAACCAAGTAAAGCTCTTAAG 1705  
 QY 548 ProGlnGlyPheLeuProProProPheProProProGlyGlyValAspArgGlyValGly 567  
 Db 1706 CCGTATGTCTCTTTT---CCTCTTTCACCTTCAAGAGAGAGAACAGTGAAGAGTGGG 1762  
 QY 568 ProGlyGlyMetArgGlyValArgGlyGlyLysMetAspArgGlyValProGlyGlyMet 587  
 Db 1763 CCC---GGCATGAGAGAGAGAGAG---GGCTTATGAGCCGTGAGAGGCTT---GGTATG 1813  
 QY 588 PheArgGlyValArgGlyValAspArgGlyGlyPheArgGlyValArgGlyMetAspArg 607  
 Db 1814 TTCAGAGAGTGGCCGTGAGAGAGATAGAGAGGAGGCTTCAG---GGCCGTGATGACAG 1870  
 QY 608 GlyValPheGlyValGlyValArgArgGlyGlyProGlyGlyProProGlyProLeuMetGlu 627  
 Db 1871 ---GATTTGTGTGAAG 1927  
 QY 628 GlnMetGlyValArgArgGlyGly-----ArgGlyGlyProGlyLysMetAspLysGly 645  
 Db 1928 CAATCTGTGTGAAGAGTGTGTGTGTGAAGAGAGAGTGTGTGTGTGTGTGTGTGTGT 1987  
 QY 646 GluHisArgGlnGlnArgArgArgAspArgProTyr 656  
 Db 1988 GACCATGCGCAG 2020  
 RESULT 15  
 AK014366 1379 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK014366  
 DEFINITION Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched library, clone:330002D11 product: Ewing sarcoma homolog, full insert sequence.  
 ACCESSION AK014366  
 VERSION AK014366.1 GI:12852167  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL 99279253  
 MEDLINE 10349636  
 PUBMED 10349636  
 REFERENCE 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komu, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL 20493374  
 MEDLINE 11042159  
 PUBMED 11042159  
 REFERENCE 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komu, H., Akiyama, J., Nishi, K., Kitsumai, T., Teshiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,



Db 783 CCTGGAGGAATGTTCAAGAGTGGCAGAGTGGAGACAGAGAGGCTTCCGAGGTGGCCGT 842  
QY 604 G1yMeCAspArg1yG1yPheG1yG1yAArgArg1yG1yProG1yG1yProProG1y 623  
Db 843 GGAATGACCGAGGTGCTTTGTGTGAGAGACGAGTGTGCTCTGGGGGCTCTCTGGA 902  
QY 624 ProLeuMeG1uG1uMeG1yG1yAArgArg1yG1yAArg1yG1yProG1yLybMeCAsp 643  
Db 903 CCTTTAATGAAACGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 962  
QY 644 LybG1yG1uH1bArgG1uArgG1uArgArgAspArgProTyr 656  
Db 963 AAAGCGAGCACCGTCAGAGAACGAGAGACCGGCTTAC 1001

Search completed: February 21, 2005, 04:11:40  
Job time : 4977 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 15:13:01 ; Search time 165 Seconds  
(without alignments)  
792.273 Million cell updates/sec

Title: US-10-791-017a-2\_COPY\_319\_656  
Perfect score: 1922  
Sequence: 1 GGRGCMGSGNGRGNKPGG.....GGRGMDKGNHGRDRRY 338

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp19808:\*  
2: geneseqp19908:\*  
3: geneseqp20008:\*  
4: geneseqp20018:\*  
5: geneseqp20028:\*  
6: geneseqp20038:\*  
7: geneseqp20038a:\*  
8: geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	600	ADA55145	ADA55145 Human pro
2	1922	100.0	656	ADA516278	ADA516278 Human Bv1
3	1922	100.0	656	ABM82330	ABM82330 Tumour-as
4	1922	100.0	665	ABG06460	ABG06460 Novel hum
5	1913	99.5	656	AAR44555	AAR44555 Human EMS
6	1909	99.3	604	ADR09402	ADR09402 Human pro
7	1900	98.9	656	AAW33813	AAW33813 Human EMS
8	1888.5	98.3	655	ABBS7126	ABBS7126 Human EMS
9	1810	94.2	583	ADP56334	ADP56334 Mouse lbc
10	1415	73.6	361	ABM80197	ABM80197 Tumour-as
11	878	45.7	525	ABG95084	ABG95084 Human tra
12	878	45.7	525	ADR14649	ADR14649 Human NF-
13	878	45.7	526	AAW78355	AAW78355 Human pro
14	878	45.7	526	ABG95081	ABG95081 Human tra
15	878	45.7	526	AD126117	AD126117 Human pro
16	878	45.7	526	ABM81732	ABM81732 Tumour-as
17	878	45.7	526	ADS88302	ADS88302 Human pro
18	878	45.7	569	AAW79339	AAW79339 Human pro
19	878	45.7	569	AAW79340	AAW79340 Human pro
20	873.5	45.4	518	AD126113	AD126113 Human pro
21	870	45.3	536	AAW78356	AAW78356 Human pro
22	799.5	41.6	260	ADJ68310	ADJ68310 Human hea
23	750	39.0	589	AD126207	AD126207 Human pro
24	750	39.0	592	ABO53050	ABO53050 Human put
25	750	39.0	592	AD126209	AD126209 Human pro

26	623	32.4	156	3	ABA44066	ABA44066 Human can
27	616.5	32.1	399	4	ABB60010	ABB60010 Drosophil
28	475	24.7	83	4	AAW21854	AAW21854 Peptide #
29	475	24.7	83	4	ABM44223	ABM44223 Peptide #
30	475	24.7	83	4	AAW38178	AAW38178 Peptide #
31	475	24.7	83	4	ABM27099	ABM27099 Protein #
32	475	24.7	83	4	AAW7960	AAW7960 Human bon
33	475	24.7	83	4	AAW65255	AAW65255 Human bra
34	475	24.7	83	4	ABG59596	ABG59596 Human liv
35	475	24.7	83	5	ABG46973	ABG46973 Human pep
36	419	21.8	266	7	ADL22735	ADL22735 Human dis
37	407.5	21.2	280	8	AD126115	AD126115 Human pro
38	303	15.8	1078	2	AAW71704	AAW71704 Collagen
39	303	15.8	1078	3	AAW96125	AAW96125 Collagen
40	303	15.8	1078	5	AAE16478	AAE16478 Human col
41	303	15.8	1078	5	ABM80736	ABM80736 Collagen
42	303	15.8	1078	5	ABM809628	ABM809628 Amino aci
43	303	15.8	1078	7	ADP13078	ADP13078 Human col
44	303	15.8	1466	4	AAW02534	AAW02534 Bovine al
45	303	15.8	1466	4	AAW02533	AAW02533 Bovine al

## ALIGNMENTS

RESULT 1  
ADA55145  
ADA55145 standard; protein; 600 AA.  
XX  
AC ADA55145;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human protein, SEQ ID 2713.  
XX  
KW Cytotoxic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KW Inflammatory disease; osteoporosis; neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN EP1293569-A2.  
PD  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
PA (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
XX WPI; 2003-395539/38.  
XX DR N-PSDB; ADA53506.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 14; SEQ ID NO 2713; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX Sequence 600 AA:

	Query Match	100.0%	Score 1922;	DB 6;	Length 600;
	Best Local Similarity	100.0%	Pred. No. 5,4e-148;		
	Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	1	GGRGGMSAGERGGFNKPGPMDEGDLIDLGPVDPDEBDSNLSAIYVGLNDSVTLDDL	60		
Db	263	GGRGGMSAGERGGFNKPGPMDEGDLIDLGPVDPDEBDSNLSAIYVGLNDSVTLDDL	3222		
Qy	61	DFFRGQGVVMNRRQTQPMIHTIYLDKERTKPKGDA TVSYEDPPTAKAA VEFHFDKDRQS	120		
Db	323	DFFRGQGVVMNRRQTQPMIHTIYLDKERTKPKGDA TVSYEDPPTAKAA VEFHFDKDRQS	3822		
Qy	121	KLKVLARKKPPMNSMRGSLPPEEGKMPPLRGGPGEFGGPGMGMRGGDRGAFP	180		
Db	383	KLKVLARKKPPMNSMRGSLPPEEGKMPPLRGGPGEFGGPGMGMRGGDRGAFP	4424		
Qy	181	PRGPRGSRGNPSGGGVVQHRADGMQCPNPGCNQNPAMRTGNCQKAPKPEGLPPPPP	240		
Db	443	PRGPRGSRGNPSGGGVVQHRADGMQCPNPGCNQNPAMRTGNCQKAPKPEGLPPPPP	5020		
Qy	241	PGGRRGRRGGGRRGRRGGLMDRGPGGMFRGGRGDGGPPRGGRGMDRGFRGGRRGCP	300		
Db	503	PGGRRGRRGGGRRGRRGGLMDRGPGGMFRGGRGDGGPPRGGRGMDRGFRGGRRGCP	5622		
Qy	301	GGPPGPLMEQMGRRGGRGGPGKMDGSHRQRRDRPY 338			
Db	563	GGPPGPLMEQMGRRGGRGGPGKMDGSHRQRRDRPY 600			
RESULT 2					
ADSI6278					
ID	ADSI6278	standard; protein; 656 AA.			
AC	ADSI6278;				
DT	18-NOV-2004	(first entry)			
DE	Human Ewing sarcoma protein Segid 2.				
XX	human; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;				
KM	hypogonadism; androgen-resistance syndrome; testicular feminisation.				
XX	Homo sapiens.				
OS	Homo sapiens.				
PN	EP1455190-A1.				
XX	08-SEP-2004.				
PD	08-SEP-2004.				
PF	16-FEB-2004; 2004EP-00003422.				
XX	04-MAR-2003; 2003DE-01009280.				
PR	25-APR-2003; 2003US-0465692P.				
XX	(SCHD ) SCHERING AG.				
PA	Obendorf M, Wolf S;				
PI	Obendorf M, Wolf S;				
XX	WPI, 2004-627861/61.				
DR	N-PSDB; ADSI6277.				
XX					
PT	Determining the hormonal effects of substances, used to identify				
PT	pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,				
PT	from modulating interaction between nuclear receptors and Ewing sarcoma				
PT	protein.				
XX					
XX	Disclosure; SEQ ID NO 2; 30pp; German.				
XX					
CC	This invention relates to a novel modulators that alter the interaction				
CC	between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well				
CC	as the screening method thereof. Specifically, it refers to determining				
CC	and identifying a hormonal effect brought about by test compounds that				
CC	modulate either the binding of EWS to the nuclear receptor or the ligand-				

Query Match	100.0%	Score 1922;	DB 8;	Length 656;
Best Local Similarity	100.0%	Pred. No. 6e-148;		
Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	GGRGMSAGRGGFENKRGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDLA	60		
319	GGRGMSAGRGGFENKRGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDLA	378		
61	DFPQCCVVMANKRTGQPMIHIYLDKETGPKGDATVSYEDPPAKAAVEMFDGKDQGS	120		
379	DFPQCCVVMANKRTGQPMIHIYLDKETGPKGDATVSYEDPPAKAAVEMFDGKDQGS	438		
121	KLKVLARKKPPMNSMNRGLPPREGMRPPLRLRGPGCGPGGPMGRMGGRGGRGCFP	180		
439	KLKVLARKKPPMNSMNRGLPPREGMRPPLRLRGPGCGPGGPMGRMGGRGGRGCFP	498		
181	PRGRSGRGNPGSGGANNVHRAGDMQCNPPCGNCFMARTBCNCKKAPKDEGFLPPFP	240		
499	PRGRSGRGNPGSGGANNVHRAGDMQCNPPCGNCFMARTBCNCKKAPKDEGFLPPFP	558		
241	PGGRGRGPGCGMKGGRGLMDRGGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRG	300		
559	PGGRGRGPGCGMKGGRGLMDRGGPGMFRGGRGGRGGRGGRGGRGGRGGRGGRG	618		
301	GGPPGLMEQNGRRGRRGGFGKMDKGEHRQERRDRY	338		
619	GGPPGLMEQNGRRGRRGGFGKMDKGEHRQERRDRY	656		
RESULT 3				
ABM82330				
ID	ABM82330 standard; protein; 656 AA.			
AC	ABM82330;			
XX				
XX				
DT	18-NOV-2004 (first entry)			
XX				
DE	Tumour-associated antigenic target (TAT) polypeptide PRO58232, SEQ:5990.			
XX				
XX				
KM	Tumour-associated antigenic target; TAT; human; overexpression; cancer;			
KM	tumour diagnosis; cell proliferative disorder; breast cancer;			
KM	colorectal cancer; lung cancer; ovarian cancer; liver cancer;			
KM	central nervous system cancer; bladder cancer; pancreatic cancer;			
KM	cervical cancer; melanoma; leukaemia; hybridisation probe;			
KM	chromosome identification; chromosome mapping; gene mapping;			
KM	gene therapy; cytostatic.			
XX				
OS	Homo sapiens.			
XX				
PN	WO2004030615-A2.			
XX				
PD	15-APR-2004.			
XX				
PF	29-SEP-2003; 2003WO-US028547.			
XX				
PR	02-OCT-2002; 2002US-0414971P.			
XX				
PA	(GETH ) GENENTECH INC.			
XX				

PI Wu TD, Zhang Z, Zhou Y;  
XX WPI, 2004-347921/32.  
DR N-PSDB; ACN40903.  
XX  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
XX Claim 12; SEQ ID NO 5990; 7273bp; English.  
XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid, an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention  
XX  
XX  
SQ Sequence 656 AA;  
Query Match 100.0%; Score 1922; DB 8; Length 656;  
Best Local Similarity 100.0%; Pred. No. 6e-148; Indels 0; Gaps 0;  
Matches 338; Conservative 0; Mismatches 0;  
QY 1 GGRGSGSAGERGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIVYQGLNDSVTLLDILA 60  
DB 319 GGRGSGSAGERGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIVYQGLNDSVTLLDILA 378  
QY 61 DFFKQCGVVMKRTQPMIHIYLDKRTGPKGDATVSYEDPPTAKAAVEMFGKDFQGS 120  
DB 379 DFFKQCGVVMKRTQPMIHIYLDKRTGPKGDATVSYEDPPTAKAAVEMFGKDFQGS 428  
QY 121 KLVSLARKKPPMNSMRGGLPPREGGMPPLRGSGPGSGPGMGRMGSGDGGFP 180  
DB 439 KLVSLARKKPPMNSMRGGLPPREGGMPPLRGSGPGSGPGMGRMGSGDGGFP 498  
QY 181 PRGPRGSRGNPSGGGVNORAGDMQCPNPGCGNFMAMTECNQCAKPEBGLPPFPFP 240  
DB 499 PRGPRGSRGNPSGGGVNORAGDMQCPNPGCGNFMAMTECNQCAKPEBGLPPFPFP 558  
QY 241 PGGDRGRGPGGMRGGRGGLMDRGPGMFRGGRGGDGGFRGGRGMDRGFGGGRGPG 300  
DB 559 PGGDRGRGPGGMRGGRGGLMDRGPGMFRGGRGGDGGFRGGRGMDRGFGGGRGPG 618  
QY 301 GGPGLMEOMGGRGGRGGGPGKMDKGEHRQBRDRPY 338  
DB 619 GGPGLMEOMGGRGGRGGGPGKMDKGEHRQBRDRPY 656  
RESULT 4  
ID ABG06460 strand; protein; 665 AA.  
XX ABG06460;  
AC  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6451.

XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI, 2001-639362/73.  
DR  
XX N-PSDB; AAS70647.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 36819; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences  
XX  
XX  
SQ Sequence 665 AA;  
Query Match 100.0%; Score 1922; DB 4; Length 665;  
Best Local Similarity 100.0%; Pred. No. 6.1e-148; Indels 0; Gaps 0;  
Matches 338; Conservative 0; Mismatches 0;  
QY 1 GGRGSGSAGERGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIVYQGLNDSVTLLDILA 60  
DB 328 GGRGSGSAGERGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIVYQGLNDSVTLLDILA 387  
QY 328 GGRGSGSAGERGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIVYQGLNDSVTLLDILA 60  
DB 388 DFFKQCGVVMKRTQPMIHIYLDKRTGPKGDATVSYEDPPTAKAAVEMFGKDFQGS 447  
QY 61 DFFKQCGVVMKRTQPMIHIYLDKRTGPKGDATVSYEDPPTAKAAVEMFGKDFQGS 120  
DB 388 DFFKQCGVVMKRTQPMIHIYLDKRTGPKGDATVSYEDPPTAKAAVEMFGKDFQGS 447  
QY 121 KLVSLARKKPPMNSMRGGLPPREGGMPPLRGSGPGSGPGMGRMGSGDGGFP 180  
DB 448 KLVSLARKKPPMNSMRGGLPPREGGMPPLRGSGPGSGPGMGRMGSGDGGFP 507  
QY 181 PRGPRGSRGNPSGGGVNORAGDMQCPNPGCGNFMAMTECNQCAKPEBGLPPFPFP 240  
DB 508 PRGPRGSRGNPSGGGVNORAGDMQCPNPGCGNFMAMTECNQCAKPEBGLPPFPFP 567

[illegible]

Db	439	KLK1SLARKKCPKPNMSMRGLPLPRREGRGMPPLRLGGPGPGGCGPMGSGRGGDRGCGFP	498
Qy	181	PRGRGRSGNGNSGGGNNVQHRAGDMQCPRPGCGNQNQFARTENOCCKAPPEBSFLPPFP	240
Db	499	PRGRGSGSGNSGGGNNVQHRAGDMQCPRPGCGNQNQFARTENOCCKAPPEBSFLPPFP	558
Qy	241	PGGDRGRGRGPGGMRGGRGGLMDRGGPGMPFRGGRGDRGGPFGRGMDRGFGGGRG	300
Db	559	PGGDRGRGRGPGGMRGGRGGLMDRGGPGMPFRGGRGDRGGPFGRGMDRGFGGGRG	618
Qy	301	GGPPGPLMEQMGRRGRGGRGPGKMDKGEHREDRDPY	338
Db	619	GGPPGPLMEQMGRRGRGGRGPGKMDKGEHREDRDPY	656
RESULT 6			
ADRO9402			
ID	ADRO9402	standard; protein; 604 AA.	
XX	ADRO9402;		
AC			
XX			
DT	04-NOV-2004	(first entry)	
DE			
XX			
XX		Human protein useful for treating neurological disease Seq 2908.	
KM		human; oligo-capping method; diagnostic marker; gene therapy;	
KM		osteoporosis; neurological disease; Alzheimer's disease;	
KM		Parkinson's disease; dementia; short memory; cancer;	
KM		sense or motor function; emotional reaction; fear response; panic;	
KW		osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;	
XX		tranquilliser.	
XX			
OS		Homo sapiens.	
PN		EPI447413-A2.	
PD		18-AUG-2004.	
PF		12-FEB-2004; 2004EP-00003145.	
PR		14-FEB-2003; 2003JP-00102207.	
PR		09-MAY-2003; 2003JP-00131452.	
XX			
PA		(REAS-) RES ASSOC BIOTECHNOLOGY.	
PI		Isegai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;	
PI		Wakamatsu A, Ishii S, Nagai K, Irie R;	
DR		WPI, 2004-583265/57.	
XX		N-PSDB; ADRO7446.	
PT			
XX			
PS		Claim 1; SEQ ID NO 2908; 2686bp; English.	
CC			
CC		This invention relates to novel, isolated full length human cDNA	
CC		molecules and the encoded proteins thereof. Specifically, it refers to	
CC		cDNA clones obtained by an oligo-capping method, where none of these	
CC		clones are identical to any known human mRNAs. The present invention	
CC		describes an immunoassay to identify agonists and antagonists, as well as	
CC		antibodies, antisense molecules and siRNAs that can all be used to bind	
CC		to and modulate expression of the cDNA molecules. As such, these	
CC		molecules are useful for diagnostic markers or therapeutic targets for	
CC		the various diseases or morbid states. In particular, they are useful in	
CC		gene therapy for treating osteoporosis, neurological disease, Alzheimer's	
CC		disease, Parkinson's disease, dementia, short memory and various cancers,	
CC		as well as for maintaining equilibrium of sense or motor function, and	
CC		for treating emotional reaction, fear response and panic. Accordingly,	
CC		they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,	
CC		cyostatic and tranquilliser activities. This polypeptide is a protein	
CC		encoded by a full length human cDNA sequence of the invention. NOTE: This	

CC sequence is not given in the sequence listing of the specification but  
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-  
CC office.

XX Sequence 604 AA;

Query Match 99.3%; Score 1909; DB 8; Length 604;  
Best Local Similarity 99.7%; Pred. No. 6.2e-147;  
Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRGMSAGERGKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 60  
DB 267 GGRGMSAGERGKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 326  
QY 61 DFFKQCGVVMNKRRTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDKDFGS 120  
DB 327 DFFKQCGVVMNKRRTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDKDFGS 386  
QY 121 KLVSLARKKPPMNSMRGGLPREGGMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGG 180  
DB 387 KLVSLARKKPPMNSMRGGLPREGGMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGG 446  
QY 181 PRGPRGSRGNPSGGGVQHRAGDMQCPNPGCGNQNFAMRTGCKAPKPEGFLPPFPFP 240  
DB 447 PRGPRGSRGNPSGGGVQHRAGDMQCPNPGCGNQNFAMRTGCKAPKPEGFLPPFPFP 506  
QY 241 PGDRGRGPGGMRGGRGGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 300  
DB 507 PGDRGRGPGGMRGGRGGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 566  
QY 301 GGPPLMEQMGRRGGRGGRGPGKMDKGEHRQERRDRPY 338  
DB 567 GGPPLMEQMGRRGGRGGRGPGKMDKGEHRQERRDRPY 604

RESULT 7  
AAW33813  
ID AAW33813 standard; protein; 656 AA.

XX AAW33813;  
XX 06-JUL-1998 (first entry)  
XX Human EMS protein.  
XX Human EMS protein.  
XX Tat stimulatory factor; Tat-SF1; transcriptional activity factor; HIV-1;  
XX Infection; gene therapy; EMS.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Region 30..44  
XX /note="Imperfect repeat peptide showing homology to Tat-SF1"  
XX Region 209..236  
XX /note="Imperfect repeat peptide showing homology to Tat-SF1"  
XX W09800695-A2.  
XX 08-JUN-1998.  
XX 03-JUL-1997; 97WO-US011713.  
XX 03-JUL-1996; 96US-0021218P.  
XX 13-DEC-1996; 96US-0033152P.  
XX (MAST) MASSACHUSETTS INST TECHNOLOGY.  
XX Sharp PA, Zhou Q;  
XX WPI; 1998-087086/08.

PT Nucleic acid encoding Tat stimulating factor protein and related  
PT transformed cells - proteins and binding agents, used to treat human  
PT immunodeficiency virus infection.

XX Example 6; Page 47-51; 68pp; English.

CC EMS is a member of a novel family of putative transcription factors that  
CC have RNA recognition motifs and which are frequently associated with many  
CC types of sarcomas. It shows homology to a novel human transcriptional  
CC activity factor, Tat-stimulatory factor (Tat-SF1, see AAW33811), that is  
CC involved in the regulation of transcriptional elongation of HIV-1 by Tat.  
CC is essential for Tat trans-activation and is a substrate of an associated  
CC cellular kinase. Tat-SF can be used to screen for binding agents useful  
CC in the treatment of HIV infection

XX Sequence 656 AA;

Query Match 98.9%; Score 1900; DB 2; Length 656;  
Best Local Similarity 99.4%; Pred. No. 3.7e-146;  
Matches 336; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGRGMSAGERGKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 60  
DB 319 GGRGMSAGERGKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 378  
QY 61 DFFKQCGVVMNKRRTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDKDFGS 120  
DB 379 DFFKQCGVVMNKRRTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDKDFGS 438  
QY 121 KLVSLARKKPPMNSMRGGLPREGGMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGG 180  
DB 439 KLVSLARKKPPMNSMRGGLPREGGMPPLRGPGGPGGPGGPGGPGGPGGPGGPGGPGG 498  
QY 181 PRGPRGSRGNPSGGGVQHRAGDMQCPNPGCGNQNFAMRTGCKAPKPEGFLPPFPFP 240  
DB 499 PRGPRGSRGNPSGGGVQHRAGDMQCPNPGCGNQNFAMRTGCKAPKPEGFLPPFPFP 558  
QY 241 PGDRGRGPGGMRGGRGGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 300  
DB 559 PGDRGRGPGGMRGGRGGLMDRGPGGMPFRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 618  
QY 301 GGPPLMEQMGRRGGRGGRGPGKMDKGEHRQERRDRPY 338  
DB 619 GGPPLMEQMGRRGGRGGRGPGKMDKGEHRQERRDRPY 656

RESULT 8  
ABB57126  
ID ABB57126 standard; protein; 655 AA.

XX ABB57126;  
XX 07-MAR-2002 (first entry)  
XX Mouse ischemic condition related protein sequence SEQ ID NO:289.  
XX Mouse; ischemia; compressive ischemia; occlusive ischemia;  
XX vasoepastic ischemia; ischemic condition; ischemic disease.  
XX Mus musculus.  
XX W0200188188-A2.  
XX 22-NOV-2001.  
XX 18-MAY-2001; 2001WO-JP004192.  
XX 18-MAY-2000; 2000JP-00145977.  
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR MPI: 2002-034733/04.  
DR N-PSDB; ABI99383.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or by  
PT determining the expression profile of a gene group comprising these  
XX genes.

PS Claim 2, Page 799-802; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (1) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (1). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the  
CC protein sequences in AB57020 to AB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischaemic condition-improving drugs or  
CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
CC primers for a mouse ischaemic condition related sequence, which are used  
CC in the exemplification of the present invention

XX Sequence 655 AA;

Query Match 98.3%; Score 1888.5; DB 5; Length 655;

Best Local Similarity 98.2%; Pred. No. 3.2e-145;

Matches 332; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 GGGGMSAGRGSGPNNKGGPMDGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA 60  
DB 319 GGGGGLG-AGEGGRFNNKGGPMDGPDLDGALPDEDESDNSAIYVQGLNDSVTLLDIA 377  
QY 61 DFFKQGVVVMNKRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVEMPDGKDFQGS 120  
DB 378 DFFKQGVVVMNKRRTGQPMIHIYLDKETGPKGDATVSYEDPPTAKAAVEMPDGKDFQGS 437  
QY 121 KLVSLARKKPPNNNRGGLPFBEGRMPPPLRGSGPFGSGPMGRMGSGGDRGGRGP 180  
DB 438 KLVSLARKKPPNNNRGGLPFBEGRMPPPLRGSGPFGSGPMGRMGSGGDRGGRGP 497  
QY 181 PRGPRGSRGNPSGGGVNQRAGMOCNPGCGNPNFAMRTGECNQCCKAPKREGLPPPPPP 240  
DB 498 PRGPRGSRGNPSGGGVNQRAGMOCNPGCGNPNFAMRTGECNQCCKAPKREGLPPPPPP 557  
QY 241 PGDGRGSGPNNKGGPMDGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA 300  
DB 558 PGDGRGSGPNNKGGPMDGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA 617  
QY 301 GGGPPLMEQMGRRGSGPNNKGGPMDGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA 338  
DB 618 GGGPPLMEQMGRRGSGPNNKGGPMDGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA 655

RESULT 9

IDP56334 standard; protein; 583 AA.

ADP56334;

18-NOV-2004 (first entry)

XX Human PRO protein sequence SEQ ID NO:2310.

XX human; PRO; immune related disease; inflammatory immune response;  
KM immune response stimulation; antiallergic; antianemic; antiathletic;  
KM antisthmatic; antidiabetic; antiinflammatory; antipruritic;  
KM antineumatic; antitumor; CNS; dermatological; gastroenteric;  
KM haemostatic; hepatotrophic; immunostimulant; immunosuppressive; muscular;  
KM nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
KM virucide; gene therapy.

XX Homo sapiens.

OS WO2004039956-A2.

PN 13-MAY-2004.

PD 28-OCT-2003; 2003WO-US034381.

XX 29-OCT-2002; 2002US-0422472P.

XX (GENT) GENENTECH INC.

PI Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;

XX Wood WT, Wu TD;

XX MPI: 2004-376182/35.

DR N-PSDB; ADP56333.

XX New PRO polynucleotides and polypeptides, useful in useful in diagnosing  
PT and treating an immune related disease, e.g. systemic lupus and in  
PT erythematous, rheumatoid arthritis, diabetes mellitus or asthma and in  
PT stimulating an immune response.

XX Claim 1; SEQ ID NO 2310; 3009pp; English.

XX The present invention describes an isolated PRO nucleic acid (1). Also  
CC described: (1) a vector comprising (1); (2) a host cell comprising the  
CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an  
CC isolated PRO polypeptide; (5) a chimeric molecule comprising the  
CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
CC antibody which specifically binds to a polypeptide of (4); (7) a  
CC composition of matter comprising a polypeptide of (4), an agonist or  
CC antagonist of the polypeptide or an antibody that binds to the  
CC polypeptide in combination with a carrier; (8) an article of manufacture  
CC comprising a container, a label on the container and a composition of  
CC matter of (7); (9) a method of treating an immune related disease in a  
CC mammal; (10) a method for determining the presence of a PRO polypeptide  
CC in a sample suspected of having the polypeptide; (11) a method of  
CC diagnosing an immune related disease or an inflammatory immune response  
CC in a mammal; (12) a method of identifying a compound that inhibits or  
CC mimics the activity of or expression of a gene encoding a PRO polypeptide  
CC; and (13) a method of stimulating the immune response in a mammal. The  
CC PRO sequences have antiallergic, antianemic, antiarthritic,  
CC antisthmatic, antidiabetic, antiinflammatory, antipruritic,  
CC antineumatic, antitumor, CNS, dermatological, gastroenteric,  
CC haemostatic, hepatotrophic, immunostimulant, immunosuppressive, muscular,  
CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and  
CC virucide activities, and can be used in gene therapy. The nucleic acid  
CC (1) and the encoded polypeptides, compositions, kits and methods are  
CC useful in diagnosing and treating an immune related disease and in  
CC stimulating an immune response. The present sequence represents a human  
CC PRO protein from the present invention.

XX Sequence 583 AA;

Query Match 94.2%; Score 1810; DB 8; Length 583;

Best Local Similarity 100.0%; Pred. No. 7e-139;

Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA DFFKQGVVVMNKRRTGQPMI 80  
DB 266 PMDEGPDLDGPPVDEDESDNSAIYVQGLNDSVTLLDIA DFFKQGVVVMNKRRTGQPMI 325  
QY 81 HIYLDKETGPKGDATVSYEDPPTAKAAVEMPDGKDFQGS KLVSLARKKPPNNNRGGL 140  
DB 326 HIYLDKETGPKGDATVSYEDPPTAKAAVEMPDGKDFQGS KLVSLARKKPPNNNRGGL 385  
QY 141 PRGPRGSRGNPSGGGVNQRAGMOCNPGCGNPNFAMRTGECNQCCKAPKREGLPPPPPP 200  
DB 386 PRGPRGSRGNPSGGGVNQRAGMOCNPGCGNPNFAMRTGECNQCCKAPKREGLPPPPPP 445  
QY 201 AGDMQCPNPGCGNPNFAMRTGECNQCCKAPKREGLPPPPPPGDDRGSGPGRGGL 260

Db	446	AGDMQCPNFGCGNANPAMRTECHQCAKPRFEGLLPPFPFPPGGDRGCGGGMGGGGL	505
Oy	261	MDRGPGGMFRGGRGGDRGGFRGGRGMDRGGFGGRRGGFGGPPGLMEQMGRRGGG	320
Db	506	MDRGPGGMFRGGRGGDRGGFRGGRGMDRGGFGGRRGGFGGPPGLMEQMGRRGGG	565
Oy	321	PGKMDGEMRORRRDPY 338	
Db	566	PGKMDGEMRORRRDPY 583	
RESULT 10			
ABM80197	ID	ABM80197 standard, protein, 361 AA.	
XX	AC	ABM80197;	
XX	DT	18-NOV-2004 (first entry)	
XX	DE	Tumour-associated antigenic target (TAT) polypeptide PRO80700, SEQ:491.	
XX	DD	Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
KW	KW	tumour; diagnosis; cell proliferative disorder; breast cancer;	
KW	KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
KW	KW	central nervous system cancer; bladder cancer; pancreatic cancer;	
KW	KW	cervical cancer; melanoma; leukemia; hybridisation probe;	
KW	KW	chromosome identification; chromosome mapping; gene mapping;	
XX	XX	gene therapy; cytostatic.	
OS	OS	Homo sapiens.	
XX	XX	WO2004030615-A2.	
XX	XX	15-APR-2004.	
XX	XX	29-SEP-2003; 2003WO-US028547.	
XX	XX	02-OCT-2002; 2002US-0414971P.	
PR	PR	(GETH ) GENENTECH INC.	
PA	PA	Wu TD, Zhang Z, Zhou Y;	
PI	PI	WPI; 2004-347921/32.	
DR	DR	N-PSDB; ACN37537.	
XX	XX	New tumor-associated antigenic target polypeptides and nucleic acids,	
PT	PT	useful in preparing a medicament for treating or detecting a	
PT	PT	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
XX	XX	prostate cancer or tumor.	
PS	PS	Claim 12; SEQ ID NO 491; 7273pp; English.	
XX	XX	The invention relates to human tumor-associated antigenic target (TAT)	
CC	CC	polypeptides, and their related nucleic acids. The TAT polypeptides are	
CC	CC	overexpressed in cancer tissues compared to normal tissues, and may thus	
CC	CC	serve as effective targets for the diagnosis and treatment of cancer in	
CC	CC	mammals. The invention also relates to nucleic acid and polypeptide	
CC	CC	sequences at least 80% identical to the TAT nucleic acids and	
CC	CC	polypeptides; expression vectors and host cells comprising a TAT nucleic	
CC	CC	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
CC	CC	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
CC	CC	TAT polypeptide; and methods and compositions for the treatment or	
CC	CC	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
CC	CC	antibodies, antagonists, binding molecules and compositions are useful	
CC	CC	for diagnosing or treating a cell proliferative disorder associated with	
CC	CC	increased TAT expression, particularly cancers such as breast cancer,	
CC	CC	colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder	
CC	CC	cancer, pancreatic cancer, cervical cancer, cancers of the central	
CC	CC	nervous system, melanoma and leukemia. TAT nucleic acids may further be	
CC	CC	used as hybridisation probes, in chromosome and gene mapping, in	
CC	CC	chromosome identification and in gene therapy. The present sequence	

```

CC      represents a TAT polypeptide of the invention
XX
SQ      Sequence 361 AA;

Query Match      73.6%; Score 1415; DB 8; Length 361;
Best Local Similarity 78.1%; Pred. No. 6,3e-107;
Matches 261; Conservative 14; Mismatches 41; Indels 18; Gaps 2;

QY      1 GGRGMSAGRGGRGFKPKGGPMDRGPDLDLGPVDPDPEDSDNSAIYYQGLNDSVTLDDLA 60
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      45 GGCGGMSAGRGQVGFNFKPGGPMDEGPDLDLGPVDPDPEDSDNSAIYYQGLNDSVTLDDLV 104
QY      61 DPFKCGGVNKKRGTGQPMHIITLDDKETGKRGKQATVSYEDPFAKAAVEMFQDKDFQGS 120
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      105 DPFKCGGVNKKRGTGQPMHIITLDDKETKRGKQATVSCBDSPPAKAAVEMFQDKDFQGS 164
QY      121 KLKVLAKRKDPNMSMRGGLPRERGRGMPPLRGGPGGPGGPMGRMGGRGGRGGRFP 180
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      165 KLKVLAKRKDPNMSMRGGLPRERGRGMPPLRGGPGGPGGPMGRMGGRGGRGGRGLP 224
QY      181 PRGPRGSGNNSGGGNNVQHRAGDMQCCPRPGCGNQNFAWRTTCNCKCAPKBEGFLPPEPPP 240
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      225 PRGPRGSGNNSGGGNNVQHRAGDMQCCPRPGCGNQNFAWRTESNCKCAKPKBEGFLPPEPPP 284
QY      241 PGCDRGRRGPGCGMRGGRGLMDRGGPGGGMFRGGRGGRGGRGGRGGRGMDRGFGGRRGPG 300
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      285 PGDHGRGPGGMMVMGRRGLMDHDHGGPGGMFRGCGGRBRSRGGCWAMTVEA----- 335
QY      301 GGPPEPLMEQNGRRRGGRGFGKMDCKBHRQERR 334
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      336 -----LVEE---DKVALGGPRTDVNNGRKKRR 360
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11	
ABG95084	
ID	ABG95084 standard; protein, 525 AA.
XX	
AC	ABG95084;
XX	
DT	04-DEC-2002 (first entry)
XX	
DE	Human translocation (12; 16)(q13, p11) protein #5.
XX	
KW	Chromosome aberration; oncogenic fusion protein; cancer;
KW	proliferative disease; cellular protein isoform; heat shock protein 90;
KW	HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;
KW	T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;
KW	acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;
KW	acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;
KW	papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;
KW	rabdomyosarcoma; synovial sarcoma; viral infection.
XX	
OS	Homo sapiens.
XX	
PN	WO200269900-A2.
XX	
PD	12-SEP-2002.
XX	
PF	01-MAR-2002; 2002WO-US006518.
XX	
PR	01-MAR-2001; 2001US-0272751P.
XX	
PA	(CONF-) CONFORMA THERAPEUTICS CORP.
XX	
PI	Fritz LC, Burrows FJ;
XX	
DR	WPI; 2002-698710/75.
XX	
DR	N-PSDB; ABS73274.
XX	
PT	Treating genetically-defined disease associated with chromosomal
PT	aberrations yielding oncogenic fusion proteins, e.g. cell proliferative
XX	diseases, involves administering an inhibitor of heat shock protein 90.



PS Disclosure; Page 228-229; 389pp; English.

XX The invention describes a method of treating genetically-defined disease

CC associated with chromosomal aberrations yielding oncogenic fusion

CC proteins (I), treating cancerous cells containing (I) in a heterogeneous

CC cell population, treating proliferative diseases associated with mutant

CC protein or cellular protein isoforms (II) dependent on heat shock protein

CC (HSP)-90, or selectively treating cells expressing (II) involving

CC administering HSP90-inhibitor. The method is useful for treating

CC genetically-defined disease with chromosomal aberration-yielding

CC oncogenic fusion protein, treating cancerous cells containing fusion

CC protein in heterogeneous cell population, treating proliferative disease

CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or

CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.

CC p53), or selectively treating cells expressing mutant protein or cellular

CC protein isoform in a patient heterozygous for (II). The method is useful

CC for treating a disease e.g. haematopoietic disorder such as T or B cell

CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,

CC or a disease characterised by a solid tumour such as papillary thyroid

CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and

CC synovial sarcoma. The method is also useful for treating viral

CC infections. This represents a protein encoded by the DNA sequence of a

CC chromosome aberration

XX

XX Sequence 525 AA;

XX

Query Match 45.7%; Score 878; DB 5; Length 525;

Best Local Similarity 52.9%; Pred. No. 5,6e-63; Indels 72; Gaps 13;

Matches 182; Conservative 36; Mismatches 54;

QY 1 GGRGNGSAGRGFGFKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNDSVTLDDL 60

DB 248 GGRGNGGS-DRGGFNKFGPRDQSGRHD-----SEGDNSDNTTIFVQGLNDSVESA 301

QY 61 DFRKQGVVMMNRKTQPMIHIYLDKRTGPKGDATVSYDPPAKAAVEMFGKDFQS 120

DB 302 DYFQIGITKTNKTQPMIHIYLDKRTGPKGDATVSYDPPAKAAVEMFGKDFQS 361

QY 121 KLVNLSARKKPPNMSRMGGLPPREGRMPPRLGSGPGSGPGPMGRMGSGGDDGGP 180

DB 362 PIVSFPATRADN--RGGNGRGGR-----RGGPMRGCGYGG-----GSGGSGRGFP 410

QY 181 PRGPRGRGNPSGGVNVHAGDMQCPNPGCNQNFAMRTGCKQCKAPKEGFLPPFP 240

DB 411 SGG-----GGGGGQQAADWCKNPFCEMNNFMRNECKQCKAPKDG-----453

QY 241 PGDRGRGRGGRGR--GGGGLMNRGGRGGRG--GGGDRGGRGGR--GMNRGGRGG 295

DB 454 PG-----GGPGSGHMGNYDDRRGGGGRGGRGGRGGRGGRGGRGGRGGRGGRG 505

QY 296 RRGPGGPGPGLMEQWGRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338

DB 506 -----GPGKMDSRGHRDRRERY 525

RESULT 12

ID ADRI4649 standard; protein; 525 AA.

XX ADRI4649;

XX

XX 21-OCT-2004 (first entry)

XX

DE Human NF-kappaB pathway-associated protein SegID50.

XX NF-kappaB pathway; antiinflammatory; cytosolic; hepatotropic; virucide;

XX antiarthritic; antirheumatic; gastrointestinal; antiasthmatic;

XX antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;

XX immunosuppressive; vulnere; gene therapy; immune disorder;

XX inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;

XX hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;

XX hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;

XX X-linked anhidrotic ectodermal dysplasia; immunodeficiency;

KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;

KW viral replication; host cell survival; evasion of immune response;

KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;

KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;

KW autoimmune disorder; hyper immune activity;

KW aberrant acute phase response; hypercongenital condition; birth defect;

KW necrotic lesion; wound; organ transplant rejection;

KW aberrant signal transduction; proliferating disorder; cancer;

KW HIV propagation; human.

XX

OS Homo sapiens.

XX

PN WO2004065577-A2.

XX

XX 05-AUG-2004.

XX

XX 13-JUN-2004; 2004WO-US000798.

XX

XX 14-JAN-2003; 2003US-0440068P.

PR 12-MAY-2003; 2003US-0469757P.

XX

PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX

PI Nadler SG, Neubauer MG, Feder JN, Carman J;

DR MPI; 2004-562168/54.

XX

DR N-PSDB; ADRI4648.

XX

PT New isolated polynucleotides and polypeptides associated with NF-kappaB

PT pathway, useful for diagnosing, treating, or preventing disorders or

PT diseases associated with NF-kappaB pathway.

XX

PS Claim 6; SEQ ID NO 650; 237pp; English.

XX

CC This invention relates to the novel association of protein sequences (and

CC the genes which encode them) to the NF-kappaB pathway. The invention may

CC be useful for the production of compounds with an antiinflammatory,

CC cytosolic, hepatotropic, virucide, antiarthritic, antirheumatic,

CC gastrointestinal-gen, antiasthmatic, antiarteriosclerotic,

CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or

CC vulnere activity or for gene therapy. The proteins and nucleotides are

CC useful for diagnosing, preventing, treating, or ameliorating conditions

CC or diseases associated with the NF-kappaB pathway. The condition is an

CC immune disorder, an inflammatory disorder, an inflammatory disorder

CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,

CC hepatic disorders, Hodgkin's lymphoma, haematopoietic tumours, hyper-IgM

CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic

CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,

CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell

CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory

CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick

CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper

CC immune activity, disorders related to aberrant acute phase responses,

CC hypercongenital conditions, birth defects, necrotic lesions, wounds,

CC organ transplant rejection, conditions related to organ transplant

CC rejection, disorders related to aberrant signal transduction,

CC proliferating disorders, cancers and HIV propagation in cells infected

CC with other viruses. The present sequence is that of a human protein which

CC is subject to the novel association with the NF-kappaB pathway of the

CC invention. Note: This sequence does not appear in the specification but

CC was obtained by the indexer from Genbank.

XX

XX

SO Sequence 525 AA;

Query Match 45.7%; Score 878; DB 8; Length 525;

Best Local Similarity 52.9%; Pred. No. 5,6e-63; Indels 72; Gaps 13;

Matches 182; Conservative 36; Mismatches 54;

QY 1 GGRGNGSAGRGFGFKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNDSVTLDDL 60

DB 248 GGRGNGGS-DRGGFNKFGPRDQSGRHD-----SEGDNSDNTTIFVQGLNDSVESA 301

QY 61 DFRKQGVVMMNRKTQPMIHIYLDKRTGPKGDATVSYDPPAKAAVEMFGKDFQS 120

```

DB 302 DYKQIGITKNTKQTPMINTLTDRBTGLKGEATVSPDPPSARAALDPGKPSGN 361
QY 121 KLVSLARKKPPMNSMRGGLPPREGRMPPPLRGPGPGPGMGRMGGRGDRGFP 180
DB 362 PIVSFATRADPN--RGGNNGRG--RGPMRGYGGG--GSGGGRGGRFP 410
QY 181 PRGPRGRNPGSGGVNQRADGMCCPNPGCGNQNPAMRTECHNOCAPKPEGLPPFP 240
DB 411 SGG-----GGGGGQORADGMKCPNPTCEMNPBSMRNCCQCAKPKPDG----- 453
QY 241 PGGDRGRGPGGMR--GRRGLMDRGPGGMRG---GRGGDRGGRGR--GMDRGGRGG 295
DB 454 PG-----GGGGSHMGNTGDDRRGGRGGYDRGGYRGGRGDRGGRGGRGGRGFP--- 505
QY 296 RRGPGGPPGRLMEQWGRRGGRGPGKMD--KGEHRQERDRPY 338
DB 506 -----GPKMDSRGEHQRDRRERY 525

RESULT 13
AAM78355
ID AAM78355 standard; protein; 526 AA.
XX
AC AAM78355;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1017.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSB-) HYSBQ INC.
XX
PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y,
PI Ma J, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW,
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAKS1488.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 3253-3254; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating

```

```

CC activity, tissue growth factor activity, immunomodulatory activity and
CC actinin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM60020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 526 AA;
XX
Query Match 45.7%; Score 878; DB 4; Length 526;
Best Local Similarity 52.9%; Pred. No. 5,6e-63;
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

QY 1 GGRGSGSAGEGGRGPNKPGGPMDEGPDLDGPPVDEDESDNSAIVQGLSDVTLDDL 60
DB 249 GGRGSGGSG--DRGGFNFPGGPPRQGSRDH-----SEQDSDNNNTIFVQGLGEVNTLESVA 302
QY 61 DPFKQGVVYKMNRTGQPMINLITDRBTGLKGEATVSPDPPSARAALDPGKPSGN 120
DB 303 DYKQIGITKNTKQTPMINTLTDRBTGLKGEATVSPDPPSARAALDPGKPSGN 362
QY 121 KLVSLARKKPPMNSMRGGLPPREGRMPPPLRGPGPGPGMGRMGGRGDRGFP 180
DB 363 PIVSFATRADPN--RGGNNGRG--RGPMRGYGGG--GSGGGRGGRFP 411
QY 181 PRGPRGRNPGSGGVNQRADGMCCPNPGCGNQNPAMRTECHNOCAPKPEGLPPFP 240
DB 412 SGG-----GGGGGQORADGMKCPNPTCEMNPBSMRNCCQCAKPKPDG----- 454
QY 241 PGGDRGRGPGGMR--GRRGLMDRGPGGMRG---GRGGDRGGRGR--GMDRGGRGG 295
DB 455 PG-----GGGGSHMGNTGDDRRGGRGGYDRGGYRGGRGDRGGRGGRGGRGFP--- 506
QY 296 RRGPGGPPGRLMEQWGRRGGRGPGKMD--KGEHRQERDRPY 338
DB 507 -----GPKMDSRGEHQRDRRERY 526

RESULT 14
ABG95081
ID ABG95081 standard; protein; 526 AA.
XX
AC ABG95081;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human translocation (12; 16)(q13; p11) protein #2.
XX
KW Chromosome aberration; oncogenic fusion protein; cancer;
KW proliferative disease; cellular protein isoform; heat shock protein 90;
KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;
KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CMV;
KW acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;
KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;
KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;
KW rhabdomyosarcoma; synovial sarcoma; viral infection.
XX
XX Homo sapiens.
XX
XX WO200269900-A2.
XX
XX 12-SEP-2002.
XX
XX 01-MAR-2002; 2002WO-US006518.
XX
XX 01-MAR-2001; 2001US-0272751P.
XX
XX (CONF-) CONFORMA THERAPEUTICS CORP.
XX
XX Fritz LC, Burrows FJ;
XX
XX WPI; 2002-698710/75.
XX
XX N-PSDB; ABG73271.

```



```

Qy      181  PRGPRGSRGNPGSGNVORAGDMOCNPGCGNONFAMRTECNOCXAPKPEGFLPPPPPP 240
Db      412  SGG-----GGGGGQORAGDMKCPNPTCEMNNTFSWRNECNOCXAPKPDG----- 454
Qy      241  PGGDRGRGGFGGMR-GGRGGLMDRGSGFGMFRG--GRGDRGGRGGR-GMDRGFGGG 295
Db      455  PG-----GGFGGSHMGNGYDDRRGGRGYDRGGYRGRGGRGGRGGRGGRGGRGGRGGRGGR 506
Qy      296  RRGGRGGRPGPLMEQMGRRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338
Db      507  -----GPGKMDSRGEHRQDRRRRPY 526

```

Search completed: February 18, 2005, 15:27:34  
 Job time : 166 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using bw model

Run on: February 18, 2005, 15:22:02 ; Search time 45 Seconds

(without alignments)  
560.697 Million cell updates/sec

Title: US-10-791-017a-2\_COPY\_319\_656

Perfect score: 1922  
Sequence: 1 GGRGCMGSGRGERGFRKPGG.....GGPRKMDKGRHQRDRRY 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	656	2	US-08-343-443B-2
2	1922	100.0	656	3	US-09-214-564A-4
3	1922	100.0	656	4	US-09-538-092-1250
4	1810	94.2	591	4	US-09-949-016-10914
5	1810	94.2	591	4	US-09-949-016-10914
6	1337	69.6	591	4	US-09-949-016-10915
7	878	45.7	526	4	US-09-538-092-1080
8	453	23.6	86	2	US-08-343-443B-7
9	303	15.8	1078	3	US-08-963-825-21
10	303	15.8	1078	3	US-09-500-811-21
11	303	15.8	1078	3	US-09-570-573-21
12	303	15.8	1078	3	US-09-548-608-21
13	301	15.7	1057	3	US-08-931-820-4
14	266.5	13.9	955	4	US-09-949-016-8369
15	265	13.8	1603	4	US-09-949-016-6136
16	264.5	13.8	1492	4	US-08-468-996-12
17	264.5	13.8	1065	1	US-08-642-255-72
18	263.5	13.7	1341	3	US-08-963-825-18
19	263.5	13.7	1341	3	US-09-500-811-18
20	263.5	13.7	1341	3	US-09-570-573-18
21	263.5	13.7	1341	3	US-09-548-608-18
22	263.5	13.7	1461	4	US-09-585-887-9
23	263.5	13.7	1461	4	US-09-585-887-9
24	263.5	13.7	1461	4	US-09-585-887-9
25	262	13.6	310	4	US-09-331-347C-21
26	262	13.6	595	3	US-09-219-849-48
27	262	13.6	595	3	US-09-219-849-50

28	262	13.6	822	3	US-09-219-849-49	Sequence 49, Appl
29	261.5	13.6	1806	4	US-09-919-497-56	Sequence 56, Appl
30	260	13.5	633	1	US-08-642-255-73	Sequence 73, Appl
31	259	13.5	1017	4	US-08-468-996-10	Sequence 10, Appl
32	259	13.5	1057	3	US-08-931-820-3	Sequence 1, Appl
33	258.5	13.4	1060	3	US-08-931-820-3	Sequence 3, Appl
34	258.5	13.4	1418	3	US-08-963-825-20	Sequence 20, Appl
35	258.5	13.4	1418	3	US-09-010-999-1	Sequence 1, Appl
36	258.5	13.4	1418	3	US-09-500-811-20	Sequence 20, Appl
37	258.5	13.4	1418	3	US-09-570-573-20	Sequence 20, Appl
38	258.5	13.4	1418	3	US-09-548-608-20	Sequence 20, Appl
39	258	13.4	689	4	US-09-949-016-11276	Sequence 11276, A
40	258	13.4	1609	4	US-09-949-016-10910	Sequence 10910, A
41	258	13.4	1670	4	US-09-949-016-5883	Sequence 5883, Ap
42	257	13.4	599	4	US-09-949-016-8890	Sequence 8890, Ap
43	253	13.2	274	4	US-09-976-594-417	Sequence 417, App
44	253	13.2	1442	2	US-08-116-650-12	Sequence 12, Appl
45	253	13.2	1442	5	PCT-US95-02251-12	Sequence 12, Appl

## ALIGNMENTS

RESULT 1  
US-08-343-443B-2  
Sequence 2, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmaret, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougascel, Beatrice  
APPLICANT: Thomas, Jessica  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-343-443B-2

```

Query Match	100.0%	Score 1932	DB 2	length 656
Best Local Similarity	100.0%	Pred. No. 1.7e-150		
Matches 338, Conservative	0	Mismatches	0	Indels 0; Gaps 0

QY 1 GGRGMSAGERGGFNKPGSGPMDEGDDLDGPPVDEDEDSNSAIYVQGLNDSVTLLDLA 60  
DB 319 GGRGMSAGERGGFNKPGSGPMDEGDDLDGPPVDEDEDSNSAIYVQGLNDSVTLLDLA 378

QY 61 DFFKCGVGNKRGGPMTHIYLDKETGKGDATVSYEDPTAKAAVENFDGKDFQS 120  
DB 379 DFFKCGVGNKRGGPMTHIYLDKETGKGDATVSYEDPTAKAAVENFDGKDFQS 438

Qy 121 KLKSLARKKP PNNSMRGSLP PREGGMP PPLRGSGGPGGPGGPMGRMGSGGDRGGFP 180  
Db 439 KLKSLARKKP PNNSMRGSLP PREGGMP PPLRGSGGPGGPGGPGGPMGRMGSGGDRGGFP 498

QY 181 PRGPRGSRGNPSSGGGNVQHRAGDWQCNPNGGNGNQNFAMRTECNCCKAPKPEGFLLPPFP 240

Db 499 PRGPRGSRGNPSSGGGNVQHRAGDWQCNPNGGNGNQNFAMRTECNCCKAPKPEGFLLPPFP 558

Qy	241	PGGDRGRGCGCMRGGCGGLMDRGGCGMPPGCGRGCGRGGFPGGCGMDRGGCFGGRRGCGP	300
Db	559	PGGDRGRGCGPGCMRGGRGGLMDRGGCGMPPGCGRGCGDRGGFPFGCGMDRGGCFGGRRGCGP	618

[illegible]

RESULT 2  
US-09-214-564A-4  
Sequence 4 Amplification ITS/09214564A

Patent No. 6150515  
GENERAL INFORMATION  
APPLICANT: Sharp, Phillip A.  
APPLICANT: Zhou, Qiang  
TITLE OF INVENTION: Tat-SF: Cofactor For Stimulation Of Transcriptional

```

; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PaatSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

```

Query Match	100.0%	Score 1922;	DB 3;	Length 656;
Best Local Similarity	100.0%;	Pred. No. 1.7e-150;		
Matches 336;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

**Oy**      1 GGRGMSAGERRGGFNKPGPDEGPDLDPVPVDPEDSDNSAIYVGGLNDSVTLLDIA 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
**Db**    319 GGRGMSAGERGGFNKPGPDEGPDLDPVPVDPEDSDNSAIYVGGLNDSVTLLDIA 378

```

Oy      61  DFEKGGVNNKNNKRTGPMIHTYLDKETGPKPGDATSYSEDPTAKAAVWPFQKDFQS 120
Db      379  DFEKGGVNNKNNKRTGPMIHTYLDKETGPKPGDATSYSEDPTAKAAVWPFQKDFQS 438

```

QY	122	KLVKVLARKKKPPMNSMRGGLPPRESRGNPPLRGPGGPGGPPMGMGARGDRLGFP	180
Db	439	KLVKVLARKKKPPMNSMRGGLPPRESRGNPPLRGPGGPGGPPMGMGARGDRLGFP	438
QY	181	PRGPRGSRGNPSGGGNVOHRAGDMQCPNPGCGNPNFAWTECNCCKAPKEGFLPPFP	240
Db	499	PRGPRGSRGNPSGGGNVOHRAGDMQCPNPGCGNPNFAWTECNCCKAPKEGFLPPFP	558
QY	241	PGGDRRGGRGPGMRRGGGLMDRGPRGGMPPGGGGRGGRGRRGGMDRGGFGGRRGGR	300
Db	559	PGGDRRGGRGPGMRRGGGLMDRGPRGGMPPGGGGRGGRGRRGGMDRGGFGGRRGGR	618
QY	301	GCGPRFLMGCGRRRGGRGPGKMDKSGHRRERRDRPY	318
Db	619	GCGPRFLMGCGRRRGGRGPGKMDKSGHRRERRDRPY	656

RESULT 3  
US-09-538-092-1250  
; Sequence 1250, Application US/09538092

```

GENERAL INFORMATION:
APPLICANT: Gluc, Loic
APPLICANT: Mansfield, Traci A.
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
FILE REFERENCE: 15966-542

```

```

1 CURRENT APPLICATION NUMBER: US/09/538,092
2
3 CURRENT FILING DATE: 2000-03-29
4
5 PRIOR APPLICATION NUMBER: 60/127,352
6
7 PRIOR FILING DATE: 1999-04-01
8
9 PRIOR APPLICATION NUMBER: 60/178,965
10
11 PRIOR FILING DATE: 2000-02-01
12
13 NUMBER OF SEQ. ID NOS.: 1387
14
15 SOFTWARE: CurataSeqFormatter Version 0.9

```

```

/ LENGTH: 656
/
/ TYPE: PRT
/
/ ORGANISM: Homo sapiens
/
/ FEATURE:
/
/ NAME/KEY: misc_feature
/
/ LOCATION: (0)..(0)
/
/ OTHER INFORMATION: Polypeptide Accession Number Q01844
US-09-538-092-1250

```

Query Match	100.0%;	Score 1922;	DB 4;	Length 656;
Best Local Similarity	100.0%;	Pred. No. 1.7e-150;		
Matches 338; Conservative	0;	Mismatches	0;	Gaps

QY 1 GGRGMSGAGRGGFNKPGGPMDEGDDLGLCPVPDPEDSDNSAIYVQGLNDSVTLDDL 60  
DB 319 GGRGMSGAGRGGFNKPGGPMDEGDDLGLCPVPDPEDSDNSAIYVQGLNDSVTLDDL 378

QY 61 DFFKCGVKKNKRTGPMIHILYLDETGKRGDATS YEDEPTAKAAVEMFGKD FQGS 120  
Db 379 DFFKCGVKKNKRTGQPMIHILYLDKETGKRGDATS YEDEPTAKAAVEMFGKD FQGS 438

Qy 121 KLKVLSTARKKP PMSNRGLP PREGGMP PPLRGSPGGGPGGPMGMMGGRGDRGGP 18

Db 439 KLKVLSTARKKP PMSNRGLP PREGGMP PPLRGSPGGGPGGPMGMMGGRGDRGGP 49

Qy 181 PRGPRGSRGNSSGGGNVQHRAGDMQCPNPGCGNQNFAMRTETCNCKAPKEGFLPPPPPP 240  
 |||||  
 Db 499 PRGPRGSRGNSSGGGNVQHRAGDMQCPNPGCGNQNFAMRTETCNCKAPKEGFLPPPPPP 550

Qy 241 PGDDRARGGPGCMGGRGGLNDRGCGMFFRGGRGGRRGFGGRRGMDRGFGGGRRGSP 300  
Db 559 PGDDRARGGPGCMGGRGGLNDRGCGMFFRGGRGGRRGFGGRRGMDRGFGGGRRGSP 610

[illegible]



```
RESULT 4
US-09-949-016-10914
; Sequence 10914, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 10914
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10914

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-141; Length 591;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 80
DB 274 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 333
QY 81 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 140
DB 334 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 393
QY 141 PPRGGMPPPLRGFGGPGGPGPMGRMGGRGDDGFPFRPRGRSGRNPSSGGVQHR 200
DB 394 PPRGGMPPPLRGFGGPGGPGPMGRMGGRGDDGFPFRPRGRSGRNPSSGGVQHR 453
QY 201 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFPGGDRGGRGPGMRGGRGL 260
DB 454 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFPGGDRGGRGPGMRGGRGL 513
QY 261 MDRGPGMGFRGGRGDRGGRGMRGDRGFGGRRGPGGPPPLMEQMGRRGGRG 320
DB 514 MDRGPGMGFRGGRGDRGGRGMRGDRGFGGRRGPGGPPPLMEQMGRRGGRG 573
QY 321 PGMDKGEHRQERRRDPY 338
DB 574 PGMDKGEHRQERRRDPY 591

RESULT 5
US-09-949-016-10915
; Sequence 10915, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FaastSeq for Windows Version 4.0
```

```
; SEQ ID NO 10915
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10915

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-141; Length 591;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 80
DB 274 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 333
QY 81 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 140
DB 334 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 393
QY 141 PPRGGMPPPLRGFGGPGGPGPMGRMGGRGDDGFPFRPRGRSGRNPSSGGVQHR 200
DB 394 PPRGGMPPPLRGFGGPGGPGPMGRMGGRGDDGFPFRPRGRSGRNPSSGGVQHR 453
QY 201 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFPGGDRGGRGPGMRGGRGL 260
DB 454 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFPGGDRGGRGPGMRGGRGL 513
QY 261 MDRGPGMGFRGGRGDRGGRGMRGDRGFGGRRGPGGPPPLMEQMGRRGGRG 320
DB 514 MDRGPGMGFRGGRGDRGGRGMRGDRGFGGRRGPGGPPPLMEQMGRRGGRG 573
QY 321 PGMDKGEHRQERRRDPY 338
DB 574 PGMDKGEHRQERRRDPY 591

RESULT 6
US-09-949-016-7825
; Sequence 7825, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 7825
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7825

Query Match
Best Local Similarity 69.6%; Score 1337; DB 4; Length 306;
Matches 246; Conservative 13; Mismatches 37; Indels 18; Gaps 2;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 80
DB 10 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLDDLADFPKQCGVVMNKRRTGQPMI 69
QY 81 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 140
DB 70 HTYLDKETGPKGDATVSYEDPPTAKAAVWEPFGKDFQSGSKLYSLARKKPPMNSMRGGL 129
QY 141 PPRGGMPPPLRGFGGPGGPGPMGRMGGRGDDGFPFRPRGRSGRNPSSGGVQHR 200
```

Db 130 PPRGGMPPPLCGGPGFSPGPGMGMGRGGRGGLPPRRPGSGRNTSGGVVQ 189  
QY 201 AGMOCPPNPGCCGNONAMRTCCNOCAKPRRGGFLPPPPPPGGDRGRGGRGGL 260  
Db 190 AGROCPNPGCCGNONAMRTESNCKAPKEGFLPPPPPPGGDRGGRGGMGGGL 249  
QY 261 MDGPGGMPGRRGGDRGGRRGMDRGGFGGRRGGPGPPPLMEOMGRRGGRG 320  
Db 250 MDGPGGMPGRRGGDRGGRRGMDRGGFGGRRGGPGPPPLMEOMGRRGGRG 291  
QY 321 PGMDKGRHRR 334  
Db 292 PRFDVTNGRKKR 305

RESULT 7  
US-09-538-092-1080  
; Sequence 1080, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormatter Version 0.9  
; SEQ ID NO 1080  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)..(0)  
; OTHER INFORMATION: Polypeptide Accession Number P35637  
US-09-538-092-1080

Query Match 45.7%; Score 878; DB 4; Length 526;  
Best Local Similarity 52.9%; Pred. No. 1.7e-64; Indels 72; Gaps 13;  
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;  
QY 1 GGRGMSAGRRGGFNRKPGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDL 60  
Db 249 GGRGMSAGRRGGFNRKPGPMDEGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDL 60  
QY 61 DFRKQGVVANKRTGQPMTHIYDKETGPKGDATVSYDDPTAKAAVWFGKDFG 120  
Db 303 DFRKQGVVANKRTGQPMTHIYDKETGPKGDATVSYDDPTAKAAVWFGKDFG 120  
QY 121 KLVSLARKKPPMNSMGGLPRRRGRGMPPLRGGPGGPGGPMGRMGGRGGRG 180  
Db 363 PIRKVSFTRADFL-RGGGNGRGRG-RGGPGRGGYGGG-DSGGGGRGGR 411  
QY 181 PRGPRGSRGNPGGNGVQHRAGDWQCNPNGCNQNFAMRTCCNOCAKPRRGGFLPPPP 240  
Db 412 SGG-DSGGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 454  
QY 241 PGDRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 295  
Db 455 PG-GRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 506  
QY 296 RRGPGGPPPLMEOMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338  
Db 507 RRGPGGPPPLMEOMGRRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 526

RESULT 8

US-08-343-443B-7  
; Sequence 7, Application US/08343443B  
; Patent No. 5968734  
; GENERAL INFORMATION:  
; APPLICANT: Aurias, Alain  
; APPLICANT: Delattre, Olivier  
; APPLICANT: Desmaze, Chantal  
; APPLICANT: Meiot, Thomas  
; APPLICANT: Peter, Martine  
; APPLICANT: Ploougaestel, Beatrice  
; APPLICANT: Thomas, Gilles  
; APPLICANT: Zucman, Jessica  
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
; NUMBER OF SEQUENCES: 129  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Weiser & Associates  
; STREET: 230 South Fifteenth Street  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: AEDIT 1.0 DOS text editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/343,443B  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR93/00494  
; FILING DATE: 19-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/06123  
; FILING DATE: 20-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weiser, Gerard J.  
; REGISTRATION NUMBER: 19,763  
; REFERENCE/DOCKET NUMBER: 989.6121P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-875-8383  
; TELEFAX: 215-875-8394  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 86 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-343-443B-7  
Query Match 23.6%; Score 453; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5e-30;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 45 IYVQGLNDSVTLDDLDFRQGVVANKRTGQPMTHIYDKETGPKGDATVSYDDPT 104  
Db 1 IYVQGLNDSVTLDDLDFRQGVVANKRTGQPMTHIYDKETGPKGDATVSYDDPT 104  
QY 105 AKAAVWFGKDFGSKLVSLARKK 130  
Db 61 AKAAVWFGKDFGSKLVSLARKK 86  
RESULT 9  
US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per

```

1 APPLICANT: Bonde, Martin
2 TITLE OF INVENTION: A Method for Assaying Collagen Fragments
3 TITLE OF INVENTION: In Body Fluids; A Test Kit and Means for Carrying Out the
4 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
5 TITLE OF INVENTION: Disorders Associated with the Metabolism of
6 NUMBER OF SEQUENCES: 21
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Darby & Darby PC
9 STREET: 805 Third Avenue
10 CITY: New York
11 STATE: New York
12 COUNTRY: USA
13 ZIP: 10022
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 OPERATING SYSTEM: IBM PC compatible
17 SOFTWARE: Patentin Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/963,825
20 FILING DATE:
21 CLASSIFICATION: 436
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US/08/187,319
24 FILING DATE: 21-JAN-1994
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Goepfer, Adde C
27 REGISTRATION NUMBER: 29,714
28 REFERENCE/DOCKET NUMBER: 4305/08701
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 212-527-7700
31 TELEFAX: 212-753-6237
32 TELEX: 236687
33 INFORMATION FOR SEQ ID NO: 21:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1078 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39 ORIGINAL SOURCE:
40 ORGANISM: Homo sapiens
41 IMMEDIATE SOURCE:
42 CLONE: COLLAGEN ALPHA 1 (III)
43 US-08-963-825-21

```

```

QY      310 OMGGRGGRGPG-KMDKGE 328
Db      .579 GMPERGGLSGPCKDKGE 598

RESULT 10
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
GENERAL INFORMATION:
APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,811
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Geogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEX: 212-753-6237
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21

Query Match          15.8%; Score 303; DB 3; Length 1078;
Best Local Similarity 30.3%; Pred. No. 1e-16;
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

QY      1 CGRGMAGAGRRGNKKRGGPMDEBBDLGPVVPDPESDNSAIYYOGLNDSTVLDDLA 60
       | : ||||| | : | : | : | : | : | : | : | : | : | : | : | : |
Db      291 GAKGPPGRGERGEGADIPGVPAKGKDGDGSFGDPGANGLPGAAGRGLGSL----- 343

QY      61 DFFQGCGVVKNKRIGQMIIHYLDKETGPKGDATVSVEDPTTAAVAEMFPDGKDFQS 120
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      344 -----RGPAGPNGIPGSKGPAGS----- 361

QY      121 KLVSVLARKKEPMNMRG--GLPPEGRCMPPLRLGPGGPGPGG-----GPMGRMG--G 171
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      362 -----RGAGRPAPPRGAAGEPRGHDVGVRGCGRMKPSPGSPGSDGKRGPPGSGSESG 414

QY      172 RGDGRGEPFRPGPRGSRG----NPSSGGNVGHRAADMOCPNPDCCGNONFAITTECNQC 226

```

Db 415 RBPFG---PSGRGPGVWGFPGPKNDGAPKNGERGGPG-GPFGPG-----459  
Qy 227 APRKRGFLPPPP---PGDGRGRGPGGMRG---GRGLMDR-----263  
Db 460 PPKNGEYGPQGPPGPTGPGDKGDTGPRPGQLGGLPTGPGPGENEKRPKGEAG 519  
Qy 264 ---GSPGMRGGRGGRG---GFRGGRGMDRGPGGRRGPGGPG---PLMB 309  
Db 520 APAPGKGDDAGAPGEGGPPGLAGAPLKGAG-PPGPGGKGAAGPPGPGAAGTPIG 578  
Qy 310 QMGRRGRGGRGPG-KMDKGE 328  
Db 579 GMPGERGGLGSPGPKDKGE 598

RESULT 11  
US-09-570-573-21  
Sequence 21, Application US/09570573  
Patent No. 634361  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF INVENTIONS: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-570-573-21

Query Match 15.8%; Score 303; DB 3; Length 1078;  
Best Local Similarity 30.3%; Pred. No. 1e-16;  
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

1 GGRGMSAGRRGGRGGRGMDGPPDLGPPVDPBDSNRAIYVQGLNDSTVTLDDLA 60

Db 291 GAKGEPGRGERGAGIDGVPAGKEDGDKGSPDGPAGNGLPGAAERGAIS-----343  
Qy 61 DFFKQGVVKNMKRTGQPMIHTYLDKENTCKPVGDAIVSEDPPTAKAAVWFMDKDFGS 120  
Db 344 -----RGPAGNGITPEEKPAGE-----361  
Qy 121 KLVSLARKKPPMNSMRG--GLPPEGRGMPPLRGPGGPGPG---GPMGRMG--G 171  
Db 362 -----RAPAPAGRGAAGRGGRDGVPCGGMGMGMPSPGPGSDGKRPFGSGGEG 414  
Qy 172 RGDRGRGPPPRGPRGSG---NPSGGNVHRAIDWOCNPBGCCNQNFMARTCNQCK 226  
Db 415 RBPFG---PSGRGPGVWGFPGPKNDGAPKNGERGGPG-GPFGPG-----459  
Qy 227 APRKRGFLPPPP---PGDGRGRGPGGMRG---GRGLMDR-----263  
Db 460 PPKNGEYGPQGPPGPTGPGDKGDTGPRPGQLGGLPTGPGPGENEKRPKGEAG 519  
Qy 264 ---GSPGMRGGRGGRG---GFRGGRGMDRGPGGRRGPGGPG---PLMB 309  
Db 520 APAPGKGDDAGAPGEGGPPGLAGAPLKGAG-PPGPGGKGAAGPPGPGAAGTPIG 578  
Qy 310 QMGRRGRGGRGPG-KMDKGE 328  
Db 579 GMPGERGGLGSPGPKDKGE 598

RESULT 12  
US-09-548-608-21  
Sequence 21, Application US/09548608  
Patent No. 635442  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF INVENTIONS: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/548,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:

ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-548-608-21

## Query Match

15.8%; Score 303; DB 3; Length 1078;  
Best Local Similarity 30.3%; Pred. No. 1e-16;  
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

QY 1 GGRGMSAGRGGFKPPGPMDEPDLGPPVDPEDSDNSAIYQGLNDVTLDDLA 60  
DB 291 GAGGEPGRGREGAGIPGVPAAGGEGDKGSPEDPGANGIPGAAGRGALGS----- 343  
QY 61 DFFKQGVVMNKRQTQPMVHIYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFQS 120  
DB 344 -----RQDAPNGIPGKGPAGE----- 361  
QY 121 KLVSLARKKPPNNMNRG--GLPREGRGMPPLRGPGGPGPG--GPMGRMG--G 171  
DB 362 -----RGAPGAPGAGGAGBPGRDVPGPGMRGMPSPGGSGDKPKPGPSQESG 414  
QY 172 RGDGRGFPFRGSRG-----NPSGGANVQHRADMGCPNPGCCGNQNFAMTECNQCK 226  
DB 415 RPPPG--PSGPGQPGVWGFPPGKNDGAPGKNGRGPG--GPDPG----- 459  
QY 227 APKBEGLPPPPPP-----PGDGRGSGPGGMRG-----GRGLMDR----- 263  
DB 460 PPKNGETGQGPPTGPGGDKDTGPPPGQGLGPGTGPPGGENKRGEPGPKGAG 519  
QY 264 --GGPGMFRGGRGDRG-----GFRGGRGMDRGFGGRGPGGPG--PLME 309  
DB 520 APAPGKGDAAGAPGREGPPGLAGAPGLRGAG--PPREGKKAAGPPGPAAGTPTGLQ 578  
QY 310 QMGRRGGRGPG--KMDKG 328  
DB 579 GMPERGGLGSPGPKDKGE 598

## RESULT 13

US-08-931-820-4  
Sequence 8369, Application US/08931820  
Patent No. 6010863  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Assay for collagen degradation  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/931,820  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: EP 96202596.1  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1057 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Collagen type III  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1055  
OTHER INFORMATION: /label= Modified

OTHER INFORMATION: /note= "Ala may be Pro"

US-08-931-820-4

## Query Match

15.7%; Score 301; DB 3; Length 1057;  
Best Local Similarity 30.5%; Pred. No. 1.5e-16;  
Matches 116; Conservative 15; Mismatches 125; Indels 124; Gaps 18;

QY 1 GGRGMSAGRGGFKPPGPMDEPDLGPPVDPEDSDNSAIYQGLNDVTLDDLA 60  
DB 290 GAGGEPGRGREGAGIPGVPAAGGEGDKGSPEDPGANGIPGAAGRGALGS----- 328  
QY 61 DFFKQGVVMNKRQTQPMVHIYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFQS 120  
DB 329 -----GLPAAERAPGR-----GPAEPNGIPGKGPAGE----- 360  
QY 121 KLVSLARKKPPNNMNRG--GLPREGRGMPPLRGPGGPGPG--GPMGRMG--G 171  
DB 361 -----RGAPGAPGAGGAGBPGRDVPGPGMRGMPSPGGSGDKPKPGPSQESG 413  
QY 172 RGDGRGFPFRGSRG-----NPSGGANVQHRADMGCPNPGCCGNQNFAMTECNQCK 226  
DB 414 RPPPG--PSGPGQPGVWGFPPGKNDGAPGKNGRGPG--GPDPG----- 458  
QY 227 APKBEGLPPPPPP-----PGDGRGSGPGGMRG-----GRGLMDR----- 263  
DB 459 PPKNGETGQGPPTGPGGDKDTGPPPGQGLGPGTGPPGGENKRGEPGPKGAG 518  
QY 264 --GGPGMFRGGRGDRG-----GFRGGRGMDRGFGGRGPGGPG--PLME 309  
DB 519 APAPGKGDAAGAPGREGPPGLAGAPGLRGAG--PPREGKKAAGPPGPAAGTPTGLQ 577  
QY 310 QMGRRGGRGPG--KMDKG 328  
DB 578 GMPERGGLGSPGPKDKGE 597

## RESULT 14

US-09-949-016-8369  
Sequence 8369, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIORITY APPLICATION NUMBER: 60/241,755  
PRIORITY FILING DATE: 2000-10-20  
PRIORITY APPLICATION NUMBER: 60/237,768  
PRIORITY FILING DATE: 2000-10-03  
PRIORITY APPLICATION NUMBER: 60/231,498  
PRIORITY FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8369  
LENGTH: 955  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-8369

Query Match  
13.9%; Score 266.5; DB 4; Length 955;  
Best Local Similarity 30.8%; Pred. No. 9.2e-14;  
Matches 123; Conservative 19; Mismatches 143; Indels 115; Gaps 22;

QY 1 GGRG-----MGSAGERGGRGPKG---PMDEGDDLGLPP-VDPDED 39  
DB 555 GSGGSPDAGGSKLPVLANIMGSGAGKGGPGGGINVQELITITSGSPNSHPSR- 613  
QY 40 SNSAIYQGLNDVTLDDLADFFKQGVVMNKRQTQPMVHIYLDKETGPKGDATVSY 99  
DB 614 -----BLKQPYSDIKQM-LVPHGLDGPFIANGFPGGGGRGPG---NQ 656

